AMENDED PROGRESS REPORT

CONTROL SYSTEM ESTIMATION AND DESIGN FOR AEROSPACE VEHICLES

by

R. T. Stefani
T. L. Williams
S. J. Yakowitz

May, 1972

Prepared under Contract NGR 03-002-115
National Aeronautics and Space Administration

ENGINEERING EXPERIMENT STATION
COLLEGE OF ENGINEERING
THE UNIVERSITY OF ARIZONA
TUCSON, ARIZONA

Amended Progress Report

CONTROL SYSTEM ESTIMATION AND DESIGN FOR AEROSPACE VEHICLES

by

R. T. Stefani
T. L. Williams
S. J. Yakowitz

May, 1972

Distribution of this report is provided in the interest of information exchange. Responsibility for the contents resides in the authors or organization that prepared it.

Prepared under Contract NGR 03-002-115

The University of Arizona
Tucson, Arizona

for

Lewis Research Center
National Aeronautics and Space Administration

Abstract

This report is concerned with the selection of an estimator which is unbiased when applied to structural parameter estimation (i.e., the estimation of a set of unknown parameters contained in a vector h relating certain states, Y_e and X_e, which are measured with uncertainty). The form of this relationship is known and follows from the structure (nature) of some process (i.e., $Y_e = X_e h$). Structural parameter estimation is differentiated from conventional parameter estimation in which Y_e is measured with uncertainty but X_e is known exactly. The parameter h may vary with time according to the difference equation $h = \phi h_o + \Gamma w$ where ϕ and Γ are known and w is a random noise term. If X_e is known exactly a weighted least squares objective function (J) is defined wherein the error vector depends on estimates of h, resulting in a conventional weighted least squares (CWLS) estimate of h.

It is shown that the CWLS estimate is biased when applied to structural parameter estimation. Two distinct approaches to bias removal are suggested: (1) change the CWLS estimator or (2) change the objective function.

Two methods are discussed with reference to the first approach. In the subtraction method, the noise statistics are used to eliminate the bias approximately. In addition, methods are suggested for estimating the noise statistics if they are unknown. Unfortunately the new estimator eliminating the bias at the cost of increasing the variance of the estimate. In the instrumental variable (IV) method, an additional measurement is taken which, if available, removes the bias.

With reference to the second approach, an augmented objective function is minimized by linearizing the partial derivatives about previous parameter estimates. The result is a linearized iterative weighted least squares (LITWELS) technique which is the major contribution of this report. The LITWELS estimator is shown to be unbiased in an asymptotic manner when the noise statistics are known. Methods of estimating unknown noise statistics are suggested. The LITWELS estimator minimizes the residuals associated with the estimate of X_e and Y_e. A simple example problem is presented and solved using the above methods. Applications are suggested with reference to adaptive control, prosthetic devices, and image enhancement.

1.0 Introduction

Figure 1 contains a block diagram of the basic system considered in this report. The composite system is interconnected and designed to control the system output given the reference input. The control law may result from an optimization criteria, root locus considerations, etc. Contained within the system is a relationship between certain system states Y_e and χ_e. The form of this relationship is known and follows from the structure (nature) of the process. The relationship may be written in two ways.

$$Y_e = H\chi_e \qquad (1)$$

$$Y_e = X_e h \qquad (2)$$

In (1), Y_e is a column vector containing the process output states (in contrast to the system output), whereas χ_e is a vector containing the process input states (in contrast to the system input). The actual parameters contained in the matrix H are unknown. In (2), all the states contained in the vector χ_e are reformed into a matrix X_e. All the unknown parameters contained in H are reformed into a vector h.

As an example, suppose (1) is

$$\begin{bmatrix} Y_{e1} \\ Y_{e2} \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & 0 & 0 \\ 0 & 0 & h_1 & h_2 \end{bmatrix} \begin{bmatrix} X_{e11} \\ X_{e12} \\ X_{e21} \\ X_{e22} \end{bmatrix}$$

Accordingly, (3) can be written in the form of (2)

$$\begin{bmatrix} Y_{e1} \\ \\ Y_{e2} \end{bmatrix} = \begin{bmatrix} X_{e11} & X_{e12} \\ \\ X_{e21} & X_{e22} \end{bmatrix} \begin{bmatrix} h_1 \\ \\ h_2 \end{bmatrix} \qquad (4)$$

The structural parameter estimation problem as discussed in this report is a logical extension of conventional weighted least squares theory and is characterized by the following which refers to (2).

1. Y_e and X_e are measured with uncertainty
2. h is estimated in a weighted least squares sense
3. h may be time varying of the form

   $$h = \Phi h_o + \Gamma w \qquad (5)$$

   where Φ and Γ are known and w is a random noise term. In this case the estimate of h is $\hat{h} = \Phi \hat{h}_o$
4. the estimate of h must be unbiased.

In conventional weighted least squares (the abbreviation CWLS will be used) the parameter estimation problem is simpler than (5) in that X_e is known exactly.

Four basic cases are considered with reference to (5).

Problem Designation	Measurement of X_e	Parameter Matrix h
I	Noisy	Time Varying
II	Noisy	Constant
III	Exact	Time Varying
IV	Exact	Constant

Problems III and IV are within the realm of CWLS theory, whereas, Problems I and II are within the realm of structural parameter estimation theory. Problems II and IV are simpler forms of Problems I and III respectively and follow by letting $\Phi = I$ and $\Gamma = 0$. In this report, Problems I and III are treated. The results may easily be applied to Problems II and IV.

1. The conventional weighted least squares (CWLS) parameter becomes a minimum covariance estimator if the weighting matrices are properly chosen (Problem III).

2. The CWLS parameter estimator is biased when applied to structural parameter estimation (Problem I).

3. If the CWLS parameter estimator is changed by the subtraction method, the bias may be removed approximately by estimating the bias using noise statistics and noisy sensor measurements (Problem I).

4. If the CWLS parameter estimator is changed by introducing a properly chosen instrumental variable (IV), the bias may be removed (Problem I).

5. If the objective function is augmented and if the resulting partial derivatives are linearized about previous parameter estimates, a linearized iterative weighted least squares (LITWELS) parameter estimator results. If the noise statistics are used properly, the estimator is unbiased in an asymptotic manner, that is, if the n^{TH} estimate is correct, then the $n+1^{ST}$ estimate is unbiased (Problem I).

6. An example problem is worked

7. Unknown noise statistics may be estimated from the residuals of the objective function.

8. Applications are suggested

2.0 Conventional Weighted Least Squares (Problem III).

The sensor equations for measurements of Y_e and X_e are

$$\begin{aligned} Y_s &= Y_e + v \\ X_s &= X_e \end{aligned} \tag{6}$$

The noise statistics are

$$\begin{aligned} E(w) &= E(v) = 0 \\ \text{Cov } w &= Q \\ \text{Cov } v &= R \end{aligned} \tag{7}$$

Subject to the equations of 5, 6, and 7 we wish to estimate h. To estimate h, a weighted least squares objective function is defined as a function of an error $\hat{e}$ depending on h_o. One estimates h by taking $\hat{h} = \Phi\hat{h}_o$.

$$\begin{aligned} J &= \hat{e}^T M\hat{e} \\ &= (Y_s - X_s\Phi\hat{h}_o)^T M(Y_s - X_s\Phi\hat{h}_o) \end{aligned} \tag{8}$$

In order to select $\hat{h}_o$ to minimize J let us differentiate J with respect to $\hat{h}_o$. If we equate the resulting expression to zero, the following value of $\hat{h}_o$ minimizes J for positive definite M.

$$\hat{h}_o = [\Phi^T X_s{}^T M X_s \Phi]^{-1} \Phi^T X_s{}^T M Y_s \tag{9}$$

With regard to the matrix M which weights the error terms of the vector $\hat{e}$, Gauss made the following statement in 1809 concerning orbital parameters.

> "If the astronomical observations and other quantities on which the computation of orbits is based were absolutely correct, the elements also, whether deduced from three or four observations would be strictly accurate (so far indeed as the motion is

> supposed to take place exactly according to the laws of Kepler) and, therefore, if other observations were used, they might be confirmed but not corrected. But, since all our measurements and observations are nothing more than approximations to the truth . . .the most probable value of the unknown quantities will be that in which the sum of the squares of the differences between the actual observed and computed values multiplied by numbers that measure the degree of precision is a minimum.

The "degree of precision" has since been defined to be the inverse of the covariance matrix of e, where

$$\begin{aligned} e &= Y_s - X_s\Phi h_o \\ &= X_e\Gamma w + v \\ \text{cov } e &= X_e\Gamma Q\Gamma^T X_e^{\ T} + R \\ M_{\text{optimal}} &= [X_e\Gamma Q\Gamma^T X_e^{\ T} + R]^{-1} \end{aligned} \tag{10}$$

Hence, for low noise values, terms of M are large (High degree of precision) and for high noise values, terms of M are low (low degree of precision).

As a result of the choice of M_{optimal}, the covariance of the estimation error becomes simply

$$E\{(\hat{h}_o - h_o)(\hat{h}_o - h_o)^T\} \quad [\Phi^T X_s^{\ T} M X_s \Phi]^{-1} \tag{11}$$

It may also be shown that (9) is an unbiased estimate of h_o, that is

$$E\{\hat{h}_o\} = h_o \tag{12}$$

Equation 11 represents the minimum covariance matrix for the estimation error related to $\hat{h}_o$ under the following conditions

Class of Estimators	Measurement Noise
weighted least squares	zero mean, finite variance
linear	white
linear and nonlinear	white Gaussian

Let us now consider the effect of uncertainity in measuring X_e upon the expected value of $\hat{h}_o$.

3.0 Bias of the CWLS Estimator For Problem I

In addition to Equations 5, 6, and 7, let us assume that sensor measurements of X_e contain a noise N

$$\begin{aligned} X_s &= X_e + N \\ E(N) &= 0 \qquad\qquad (13) \\ E(NN^T) &= S_1 \end{aligned}$$

The sensor equation for Y_s can be written

$$Y_s = (X_s - N)\Phi h_o + X_e \Gamma w + v \qquad (14)$$

If (14) is substituted into (9) and if the expected value is taken, the following is the result where the noise sequences w, v, and N are assumed to be uncorrelated

$$\begin{aligned} E\{\hat{h}_o\} &= [I - T]\, h_o \\ T &= E\{[\Phi^T X_s^T M X_s \Phi]^{-1} \;\; \Phi^T X_s^T M N \Phi\} \end{aligned} \qquad (15)$$

We conclude that when N = 0 (CWLS parameter estimation) the result is unbiased but when $N \neq 0$ (structural parameter estimation) the result is biased and the bias must somehow be removed.

4.0 The Subtraction Method (Problem I)

One obvious method suggests itself for removing the bias present in (15): premultiply the estimator by $[I-T]^{-1}$. The expected value of $\hat{h}_o$ would then become

$$E\{\hat{h}_o\} = [I-T]^{-1}[I-T]\, h_o = h_o \qquad (16)$$

which is an unbiased estimate of h_o.

Note that for Problem I, T (the bias) depends on Φ, M, X_e and the noise statistics. We must therefore know the deterministic signal X_e, however, the entire bias problem is caused by noisy measurements of X_e, referred to as X_s. Hence, any practical bias removal method involving the matrix T requires the estimation of T (i.e. $\hat{T}$) as a function of the noise statistics and X_s. The net result is that the bias is approximately but not completely removed using $\hat{T}$. Suppose that $\hat{T}$ is chosen from (15).

$$\begin{aligned} \hat{T} &= [\Phi^T X_s{}^T M X_s \Phi]^{-1} \, \Phi^T \, \bar{T}\Phi \\ \bar{T} &= E\{X_s{}^T MN\} = E\{N^T MN\} \end{aligned} \qquad (17)$$

The following estimator results from premultiplying (9) by $[I-\hat{T}]^{-1}$ where $\hat{T}$ is defined in (17)

$$\hat{h}_o = [\Phi^T X_s{}^T h X_s \Phi - \Phi^T \bar{T}\Phi]^{-1} \, \Phi^T X_s{}^T M Y_s \qquad (18)$$

Note that the term $\Phi^T \, \bar{T}\Phi$ attempts to "subtract off" the effect of the bias, hence the term "subtraction method."

Although the bias is approximately removed by the subtraction method, there are four objections to consider. (1) The covariance of the estimation error is increased as compared to the biased CWLS estimator. (2) The above matrices contain k sets of data for multistage processes. Subtracting the probabilistic means and variances included in T when k is small may result in grossly inaccurate results since the sample means and variances may be quite different from their probabilistic counterparts. As more samples are included in the above matrices, then this method becomes more useful since the sample means and variances approach the true means and variances. (3) The statistics may be unknown, hence procedures must be sought for estimating the necessary statistics. (4) Selection of the weighting matrices is not at all well defined as in Problems III (Equation 10).

5.0 The Instrumental Variable Method (Problem I)

Let us now consider the use of an additional variable whose purpose is to achieve an unbiased estimator for Problem I. This method is called the instrumental variable (IV) method since the additional variable is an instrument for the desired result[18,19,20].

For Problem I one simply rewrites the CWLS algorithm (9) to include the instrumental variable (Z). The instrumental variable replaces X_s in such a way that the resulting estimate of h_o is unbiased.

$$\hat{h}_o = [\phi^T Z^T M X_s \phi]^{-1} [\phi^T Z^T] \, M Y_s \qquad (19)$$

The instrumental variable must be uncorrelated with the noise terms N, v, and w. If Z is thusly chosen, then substituting (14) into (19) and taking the expected value results in the proof that $E\{\hat{h}_o\} = h_o$, that is, we have an unbiased estimate of h_o.

One advantage of the IV method is that no statistics need be known to remove the bias. It is necessary, however, to find ways of choosing an instrumental variable [18,19] or forcing some variable to approach X_e[20]. The paper by Young[20] describes an interesting hybrid (analog and digital) scheme. The end result is that Z approaches X_e by proper adjustment of the model parameters. If Z is highly correlated with X_e, then the weighting matrix M is chosen as suggest (10), i.e., $M = [Z\Gamma Q\Gamma^T Z^T + R]$.

In summary, an instrumental variable Z is used as per (19). If Z is uncorrelated with the noise terms, the resultant estimator is unbiased at the expense of the covariance matrix for the estimation error. If Z is highly correlated with Z_e, the covariance matrix for the estimation error may be reduced by selecting M as above.

6.0 The Linearized Iterative Weighted Least Squares Technique (Problem I)

Let us consider a new approach for achieving unbiased structural parameter estimates. This approach follows directly from a weighted least squares minimization problem. It may be shown that unbiased results follow from a specific selection of the weighting matrices, similar to the weighting matrix selection of (10). Recall that the basic linear relationship between the states and parameters can be written in two ways

$$Y_e = X_e h \tag{20}$$

$$= H\chi_e \tag{21}$$

In view of (21) let us write the sensor equation for χ_s

$$\begin{aligned} \chi_s &= \chi_e + n \\ E(n) &= 0 \qquad (22) \\ \text{cov } n &= S \end{aligned}$$

In (5) the time variation of the parameter vector h was defined. In a similar manner, the time variation of the parameter matrix H may be defined

$$H = H_o\theta + WD \tag{23}$$

where W is a random noise term. Subject to Equations 5, 6, 7, 13, 22, and 23, we wish to estimate h. To estimate h, a weighted least squares objective function J is defined by two equivalent expressions. The first term of each expression for J depends on estimates of the noise sequence n while the second term depends on estimates of both n and h_o. One estimates h by taking $\hat{h} = \Phi\hat{h}_o$.

$$
\begin{aligned}
J &= \hat{n}^T M_1 \hat{n} + \hat{y}^T M \hat{y} \\
J &= \hat{n}^T M_1 \hat{n} + [Y_s - (Y_s - \hat{N})\Phi\hat{h}_o]^T M[Y_s - (X_s - \hat{N})\Phi\hat{h}_o] \qquad (24)\\
&= \hat{n}^T M_1 \hat{n} + [Y_s - \hat{H}_o\theta(\chi_s - \hat{n})]^T M[Y_s - \hat{H}_o\theta(\chi_s - \hat{n})]
\end{aligned}
$$

The first expression above may be differentiated with respect to $\hat{h}_o$ while the second (equivalent) expression may be differentiated with respect to $\hat{n}$. At this point, the partial derivates $\frac{\partial J}{\partial \hat{h}_o}$ and $\frac{\partial J}{\partial \hat{n}}$ form a set of nonlinear coupled equations since $\hat{N}$ depends on $\hat{n}$ and $\hat{H}_o$ depends on $\hat{h}_o$. It is possible to linearize the partial derivatives and decouple the variable $\hat{n}$ from $\hat{h}_o$ by assuming that $\hat{N}$ and $\hat{H}_o$ are constants determined by the last estimates of the variables $\hat{n}$ and $\hat{h}_o$.

The derivation of the solution to (24) begins by selecting the n^{TH} estimate of n to obtain $\hat{N}_n$. It is useful to define

$$\overline{X}_s = X_s - \hat{N}_n \qquad (25)$$

The next step is to differentiate (24) with respect to $\hat{h}_o$ to obtain an equation not involving $\hat{n}$. The solution of this equation for the $n+1^{st}$ estimate of $\hat{h}_o$ follows

$$
\begin{aligned}
\hat{h}_{o(n+1)} &= [\phi^T \overline{X}^T M \overline{X}_s \phi]^{-1}[\phi^T \overline{X}^T{}_s]MY_s \\
\hat{h}_{n+1} &= \phi\hat{h}_{o(n+1)}
\end{aligned} \qquad (26)
$$

Having a value of $\hat{h}_{o(n+1)}$ we immediately use that to obtain $\hat{H}_o$. By differentiating J with respect to $\hat{n}$ we obtain an equation independent of $\hat{h}_o$ which may be solved for the $n+1^{st}$ estimate of n

$$\hat{n}_{n+1} = [M_1 + \theta^T \hat{H}_o^T M \hat{H}_o \theta]^{-1}[\theta^T \hat{H}_o^T M \hat{H}_o \, \theta\chi_s - \theta^T \hat{H}_o^T M Y_s] \qquad (27)$$

The algorithm of (25) - (27) is therefore iterative and begins with an initial estimate of N. If the initial estimate is zero the first estimate of h is identical to the CWLS estimate.

In summary we have developed a linearized iterative weighted least square algorithm, which may be abbreviated LITWELS. This method is both iterative (requiring a fixed amount of data storage) and recursive. In contrast, the CWLS algorithm is only recursive (that is, given a fixed amount of data, only one optimal filtered result occurs).

The following procedure may be used to demonstrate that the algorithm of (25) - (27) converges in expected value to h_o in an asymptotic manner: assume that $\hat{h}_{o(n+1)} = h_o$ and show that the expected value of $\hat{h}_{o(n+2)}$ is h_o. To do so, however, requires substituting the noise estimate $\hat{n}_{n+1}$ into the parameter estimate $\hat{h}_{o(n+2)}$ which requires some way of relating $\hat{n}$ to $\hat{N}$ and $\hat{H}_o$ to $\hat{h}_o$. These relationships are difficult to establish in general since elements of the column vectors must be arranged into matrices. It must be stressed that the LITWELS technique may converge in general but that proving so is quite difficult. Convergence can be established for problems with a scalar output (i.e., each of the sets of data contained in Y_s satisfy a single equation linearly relating one output state Y_{ei} to n other states X_{ei} and n parameters h_i). Hence, for the i^{TH} set of data

$$\begin{aligned} Y_{ei} &= X_{ei}h_i \\ &= h_i^T X_{ei}^T \qquad (28) \\ &= H_i X_{ei} \end{aligned}$$

The third expression in (28) follows since X_{ei}^T is a column vector (as is X_{ei}) and h_i^T is a row vector (as is H_i when Y_{ei} is a scalar). Since X_e and χ_e may be related, then the sensor noise terms N and n may be related. Similarly the parameters H_o and h_o may be related. It is now possible to demonstrate the convergence of the LITWELS algorithm as follows

1) Let $\hat{h}_{o(n+1)} = \hat{h}_o$ so that $\hat{H}_o = H_o$. Calculate $\hat{n}_{n+1}$
2) Substitute $\hat{N}_{n+1}$ into (25) to obtain $\hat{h}_{o(n+2)}$
3) Take the expected value of the result

The resulting expected value is of the form $E\{A^{-1}B\}$ where A and B are matrices. Expected values of this type are quite difficult to evaluate, even when the probability density of the noise terms are known. Let us assume, however, that A and B contain relatively large amounts of signal and relatively small amounts of noise. Hence, A^{-1} and B are correlated only through the noise terms. If the higher moments of the noise terms are negligible (as when the noise has a Gaussian amplitude), then one can approximate the desired expected value by $E\{A^{-1}\}E(B)$ since A^{-1} and B are nearly uncorrelated. If this is done, then $E\{\hat{h}_{o_{n+2}}\} = \hat{h}_o$ when (as per (10))

$$M = [\,\overline{X}_s \Gamma Q \Gamma \overline{X}_s^{\,T} + R]^{-1}$$
$$M_1 = S^{-1} \qquad (28)$$

Note that complete knowledge of X_e implies that S approaches zero (little sensor noise is present), hence M_1 approaches zero for all iterations, hence (25) becomes the optimal conventional weighted least squares estimate as we would expect.

The above asymptotic convergence property of the LITWELS estimator, is a rather weak property. It may be possible to establish the following,

1. If the expected value of $\hat{h}_{o(n+1)}$ equals the expected value of $\hat{h}_{on}$, the expected value must be h_o.

2. The expected value of the estimator converges uniformly to h_o. This phenomenon was observed in an example problem.in which the mean of the LITWELS algorithm converged exponentially to the correct parameter value.

The storage requirements of the LITWELS algorithm is of interest. When Y_s contains scalar measurements and h is a constant parameter vector of dimension n+1, the CWLS estimator of (9) requires the storage of $\frac{n(n+3)}{2}$ running sums. It is possible to combine (25) - (27) into one equation for the constant parameter case. As a result, only one additional running sum must be stored for the LITWELS estimator. The data storage requirement is therefore minimal for the LITWELS estimator.

7.0 An Example Problem

The analysis presented in sections 1-7 appears to treat a single stage process with one set of measurements. However, the matrices Y_e, X_e, h, etc. may be defined to contain k submatrices, each one defined for one stage, that is, the equation $Y_e = X_{eh}$ may be defined to contain k equations of the form $Y_{ei} = X_{ei}h_i$. Hence, the analysis is also valid for a multistage process. For example, (9) may be written in terms of a k stage process where , X_s, etc., each contain k components.

$$\hat{h}_o = [\Sigma\Phi^T X_{si}^T M_i X_{si}^T \Phi_i]^{-1}[\Sigma\Phi_i X_{si}^T M_i Y_{si}] \qquad (29)$$

In the following multistage example problem, the parameter is constant, hence $\Phi_i = I$ and $\Gamma_i = 0$ for all stages of the process. The parameter estimation equations may be simplified accordingly and written using summations similar to (29).

Suppose we are in an automobile traveling at a constant, but unknown, velocity. The distance y_i at time t_i is related to the time t_i by the constant velocity v, that is

$$y_i = t_i \, v \tag{30}$$

In order to estimate the velocity, distance measurements are taken by reading the mileposts at five second intervals. Suppose, however, that a random error exists in the spacing of the mileposts. Suppose, in addition to the random distance measurement error, the clock has a random error associated with its timing, hence the distance measurements may be taken at actual time intervals of 4.95 sec, 5.10 sec, 5.15 sec, 4.9 sec, etc., even though the clock shows elapsed time intervals of 5.0 seconds. This type of parameter estimation problem is termed structural parameter estimation since the actual structure (parameter matrix) of the process relates quantities which are known with uncertainity.

Let us define y_{mi} as the measurement of y_i and let us define t_{mi} as the measurement of t_i. Now suppose we take k sets of measurements. A weighted least squares fit to a plot of y_{mi} versus t_{mi} may be used to estimate the constant velocity. The weighted least squares objective function J corresponding to this problem follows from (8)

$$J = \Sigma M_i (y_{mi} - t_{mi} \, \hat{v})^2 \tag{31}$$

Let us assume that each of the error terms is weighted equally (all the M_i are equal). The CWLS value of $\hat{v}$ which minimizes J follows from (9) or (29) where the M_i cancel.

$$\hat{v} = \frac{\Sigma t_{mi} \, y_{mi}}{\Sigma t_{mi}^2} \tag{32}$$

From (15) we know that the expected value of $\hat{v}$ is $(1-T)v$ where T may be defined from the random clock error N_{ti}

$$T = E\{\frac{\Sigma t_{mi} N_{ti}}{\Sigma t_{mi}^2}\} \qquad (33)$$

In order to remove the bias due to T, let us estimate T as per (17) where the variance of each noise N_{ti} is σ^2.

$$\hat{T} = \frac{k\sigma^2}{\Sigma t_{mi}^2} \qquad (34)$$

Using this estimate of T, the subtraction method estimator (18) follows by premultiplying (32) by $1/(1-T)$.

$$\hat{v} = \frac{\Sigma t_{mi} y_{mi}}{\Sigma t_{mi}^2 - k\sigma^2} \qquad (35)$$

The term $k\sigma^2$ "subtracts off" the bias due to the noisy clock readings.

Let us now consider the instrumental variable method where an additional measurement is used. For this case, the measurement must be uncorrelated with the random distance and clock errors. After each milepost reading suppose we write down the last digit of the license number of the next passing car. As an instrumental variable z_i for the i^{th} reading, let use use i times the selected digit. Since the digit is uniformly distributed in the interval 1-10, the mean value of z_i is 5i, which is also the mean value of the i^{th} time measurement (since the readings are taken at 5 second intervals). Although z_i is uncorrelated with the random time and milepost placement errors, it is also uncorrelated with t_i, which degrades the covariance of the estimation error [illegible]. The IV [illegible] estimation error but removes the bias. The IV estimator, as applied to this example, follows from (19).

$$\hat{v} = \frac{\Sigma z_i y_{mi}}{\Sigma z_i t_{mi}} \qquad (36)$$

Let us consider the LITWELS appraoch to this problem. The objective function (24) is as follows where all the M_{ii} are equal to M_i and all the M_i are equal to m

$$J = \Sigma \hat{n}^2_{ti} m_i + \Sigma[y_{mi} - (t_{mi} - \hat{n}_{ti})\hat{v}]^2 m \qquad (37)$$

If we have an n^{TH} estimate of the clock measurement errors, then the $n+1^{st}$ estimate of v follows from (25) and (26)

$$\bar{t}_{mi} = t_{mi} - \hat{n}_{ti_n}$$

$$\hat{v}_{n+1} = \frac{\Sigma \bar{t}_{mi} y_{mi}}{\Sigma \bar{t}^2_{mi}} \qquad (38)$$

With the above velocity estimate we can select on $n+1^{st}$ estimate of each measurement error from (27)

$$\hat{\hat{n}}_{ti_n} = \frac{\hat{v}^2_{n+1} m t_{mi} - \hat{v}_{n+1} m y_{mi}}{m_i + m\hat{v}^2_{n+1}} \qquad (39)$$

If σ^2 is the variance of each time measurement error, then m_i should be $1/\sigma^2$. If σ_y^2 is the variance of each distance measurement error, then m should be $1/\sigma_y^2$. Equations 38 and 39 can be combined, resulting in the LITWELS estimator for the unknown velocity problem.

$$\hat{v}_{n+2} = \frac{\{\frac{\sigma_y^2}{\sigma^2} + \hat{v}^2_{n+1}\}\ \{\frac{\sigma_y^2}{\sigma^2}\Sigma t_{mi} y_{mi} + \hat{v}_{n+1}\Sigma y^2_{mi}\}}{\frac{\sigma_y^4}{\sigma}\Sigma t^2_{mi} + 2\frac{\sigma_y^2}{\sigma^2}\hat{v}_{n+1}\Sigma t_{mi} y_{mi} + \hat{v}_{n+1}\Sigma y^2_{mi}} \qquad (40)$$

In comparison with the CWLS estimator of (32), only one additional running sum must be stored, that is, the summation of the terms y_{mi}^2.

It is important to note if the time measurement errors are small, σ^2 approaches zero, hence the subtraction estimator (35) and the LITWELS estimator (40) both approach the CWLS estimator (32). For some case, as the instrumental variable z_i approaches time values t_i the IV estimator (36) approaches the CWLS estimator. In summary, all three estimators reduce to the CWLS estimator when time is correctly measured.

Figure 2 contains the results of a simulation concerning a constant parameter problem similar to the above but of a higher dimension (2 parameters were unknown). Plotted are the sample means for 20 repetitions of each estimator.

The preceding theory is verified by Figure 2. The CWLS estimator is obviously biased from the correct value. The subtraction method results in a less biased estimator. The IV estimator is unbiased. A perfect instrumental variable was chosen. For the unknown velocity example, a perfect IV may be defined as a variable z_i equal to the correct time value t_i. A practical case where z_i must be generated would result in a power performance. The LITWEL estimator's sample mean converges toward the correct parameter value and, in fact, does so exponentially.

The above remarks support the conclusion that the four types of estimators can be ranked as follows in ascending order of their success in providing unbiased estimates with a low estimation variance

1. LITWELS
2. IV
3. Subtraction
4. CWLS

8.0 Unknown Covariance Matrices

We have considered three methods for achieving unbiased estimates with regard to the structural parameter estimation problem. The instrumental variable method requires no knowledge of the noise statistics. However, both the subtraction method and the linearized iterative weighted least squares (LITWELS) method require knowledge of certain covariance matrices. This section considers the problem of estimating the covariance matrices when the statistics are unknown. Residuals (errors and squared errors) are unknown.

For Problem I it is necessary to estimate the covariance matrices Q, R, and S which are defined for the noise vectors w, v, and n respectively. Recall that each noise vector may correspond to a k stage process so that w may contain k terms of dimension of $k_w \times 1$, v may contain k terms of dimension $k_v \times 1$, and n may contain k terms of dimension $k_n \times 1$. Suppose that the covariance matrices Q, R, and S are all nonsymmetric, that is, the noise sequences are correlated and nonstationary. There must be enough equations to estimate all the elements of Q, R, and S, hence $(kk_w)^2 + (kk_v)^2 + (kk_n)^2$ equations would be required. The resulting large number of equations would be most difficult to solve. Suppose that the covariance matrices Q, R, and S are diagonal, that is, the noise sequences are uncorrelated but not necessarily stationary. For this case, one would be required to generate $kk_1 + kk_2 + kk_3$ equations to estimate the diagonal elements of Q, R, and S. There would still be a rather large number of equations involved since k represents the number of stages. Suppose that the noise sequences are uncorrelated and stationary. In this case, only $k_1 + k_2 + k_3$ equations would be required since k_1, k_2, and k_3 elements along the main diagonals of Q, R, and S would be unique. For the purposes of this discussion, one additional simplification is used to reduce the complexity of the analysis such that only inner products (scalars) need be evaluated. It is assumed that the ratio of the elements of Q, R, and S are known, hence one need only determine scalar multiplying factors q, r, and s in order to completely specify all the matrix elements of Q, R, and S. As a result, only three scalar equations must be generated if we assume that

1. The noise vectors are uncorrelated and stationary.
2. The covariance ratios are known.

In order to generate two of the three scalar equations, one may utilize a modified CWLS estimator in which Q and R are set equal to identity matrices in (10) and the resulting value of M is used in (9). For the first equations, the sum of the residuals $\hat{e}$ from (8) may be equated to the expected value of the sum (which depends on s, h_o, Φ, Γ, and X_e). Since X_s, not X_e is available, the following approximation to this relation is made

$$\Sigma\hat{e} \approx f_1\ (s, h_o, \Phi, \Gamma, X_s) \qquad (41)$$

For a third equation, J may be evaluated for a CWLS estimate where M in (8) and (9) is an identity matrix. The result (let us use J_1 to specify the new objective function) is set equal to an approximate expression for its expected value

$$J_1 = f_3 \, (s, r, q, h_o, \Phi, \Gamma, X_s)$$

In summary, we now have three equations in four unknowns (s, r, q, and h_o).

Using the above three equations, let us consider how the subtraction method and the LITWELS method may be adjusted for the case where the noise covariances are unknown.

For the subtraction method, recall that we need an estimate of the bias matrix T which estimate depends on s, Φ, Γ, and X_s. The modified CWLS estimate may be set equal to an approximate expression for its expected value

$$\hat{h}_o \approx [I - T \, (s, \Phi, \Gamma, X_s)] \, h_o \qquad (43)$$

Equation (41) may be solved for s as a function of the unknown h_o and the result may be substituted into (43). A Newton Raphson approach is then necessary to obtain h_o from (43) since (43) is a nonlinear function of h_o.

For the LITWELS method it is necessary to use all three covariance matrices hence all three scalars must be obtained. The following scheme is suggested

1. Set h_o equal to the latest LITWELS estimate. Start with the modified CWLS estimate.

2. Solve (41) for s using h_o from 1.

3. Solve (42) and (43) for r and q. Use s from 2 and h_o from 1. Use a Newton Raphson approach

4. Obtain a LITWELS estimate using the covariance matrices resulting from 2 and 3.

5. Repeat the above steps.

The general method suggested in this section may be applied to the example problem reported in Figure 2 wherein the noise statistics are assumed to be known for a 2 dimensional unknown constant parameter problem and estimates of one of the parameters are plotted. The unknown statistics case is shown in Figure 3. The CWLS and IV estimators which do not use the noise statistics for bias removal are repeated here for reference. Note that the subtraction method estimator is less biased in Figure 3 than in Figure 2. This occurs because the subtraction method is highly dependent on the accuracy of the noise statistics and, as shown in Figure 4, the sample variance of one noise term is somewhat different than the constant $(\sigma_\delta{}^2)$ used in adjusting the CWLS estimator. For the unknown statistics case, the subtraction method procedure is able to identify the sample variance of the noise as shown in Figure 4, hence the resulting estimates are less biased as shown in Figure 3. The LITWELS method is degraded somewhat in its asymptotic appraoch to the correct parameter value. For the constant parameter problem, it is necessary for the LITWELS method to estimate r and s whereas the subtraction method estimates only s. The resulting performance is apparently degraded by this requirement. Performance, however, is quite satisfactory.

9.0 Applications

Structural parameter estimation has been discussed analytically, an example problem has been worked, and simulation results have been presented. Parameter estimation problems which can be cast into the form (5) can be solved using the methods of this report. In general, if we are concerned with a process $Y_e = X_e h$ where both Y_e and X_e are measured with uncertainity and where the time variation of h may be modeled as in (5), structural parameter estimation techniques are applicable. In this section, problems related to adaptive control, prosthetic devices, and image enhancement are discussed in relation to the structural parameter estimation.

9.1 Adaptive Control

An adaptive control system has two inter-related goals: (1) estimate the system parameters, and (2) derive a control law, using the parameters, which causes the closed loop system to perform in some desirable manner[26]. The system "adapts" to changes in its structure by adjusting the control law in an appropriate manner. Figure 1 may be related to an adaptive control system for the structural parameter estimation problem by adding a disturbance at the input and by using the parameter estimates to generate the control law . If it is possible to describe the corresponding parameter variations by (5) and if noisy measurements are taken of the states (Y_e and X_e) related by h, then the adaptive control-parameter estimation problem becomes one of structural parameter estimation. One may estimate h_o using the methods of this dissertation and then utilize $\hat{h} = \Phi\hat{h}_o$ to generate an appropriate control law. Let us consider applications which call for the use of adaptive control

(1) The handling characteristics of a vertical or short take off and landing (VSTOL) aircraft on take off are similar to that of a helicopter, but, in level flight, the handling characteristics are similar to that of a conventional airplane. It is desirable to generate a control system such that handling characteristics are constant.

(2) A shuttlecraft system has been proposed as an economical way of transporting men and material to and from space stations[28,29]. Each boost vehicle is to "fly" back to earth for future use. The shuttlecraft which rides "piggy-back" with the booster is to be highly maneuverable such that its occupants may land at certain pre-selected locations. Each suttlecraft (somewhat like a flying bathtub) must maneuver with precision since the earth's atmosphere is reentered at a very low angle and since the landing speed is in excess of 100 miles per hour. The aerodynamic conditions under which the shuttlecraft must operate are indeed varied, since the craft must fly from a near vacuum down to sea level causing the handling characteristics to vary accordingly.

(3) There are cases where unmanned spacecraft might utilize adaptive control. For example, spacecraft attempting unmanned landings on Mars and Venus encounter large aerodynamic variations. American Venus probes have indicated that surface atmospheric pressure is 75 to 100 times that of Earth[30].

9.2 Prosthetic Devices

Recent medical research has utilized the millivoltages generated by muscular contractions (myoelectric voltages) to actuate artificial limbs. The patient learns to use certain muscles to manipulate the limb.

A law governing artificial limb movement can be defined as a solution to a structural parameter estimation problem. Suppose the vector Y_s is defined to be sensed myoelectric voltages. Let the constant parameter vector be defined to be the three desired torques (t) and the three desired translational forces (f) at the time when the voltage measurements are taken.

$$h_o = \begin{bmatrix} f \\ t \end{bmatrix} \qquad (44)$$

Suppose we define a connection matrix X_s which relates the myoelectric voltages (muscle contractions) to the six degrees of freedom (h_o). A random error vector e can be defined as

$$e = Y_s - X_s\hat{h}_o \qquad (45)$$

The voltage measurements (Y_s) contain a random error. The connection matrix (X_s) is an approximation to the true relationship between the muscle contractions and the desired movement of the limb. Since we have satisfied the conditions for structural parameter estimation, $\hat{h}_o$ in (9) may be a biased estimate of the desired limb movement, making it necessary to use the methods of this dissertation to remove the bias.

9.3 Image Enhancement

With regard to received radio signals, filtering methods are used for removal of interference due to transmission of the signal. The received signal, a time function, contains the desired transmitted signal plus a noise term (the interference) which is strong in certain finite frequency bands.

An analogous problem exists when a picture is transmitted through space. The Mariner 4 took pictures of the planet Mars and reduced each image to a grid of numbers which were encoded and transmitted to Earth where the signals were decoded. The resulting data can be treated as a function of distance (by recording the numbers at intervals on a sheet of paper). The string is a sensed function $y_s(x)$

$$y_s(x) = y(x) + s(x) \tag{46}$$

where $y_s(x)$ contains the desired signal $y(x)$ plus a coherent noise term $s(x)$ which is strong in certain frequencies. Current techniques remove the coherent noise by using convolution filtering in the spatial Fourier Transform domain. Conventional least squares parameter estimation can be used as an alternate scheme.

Let us assume that the frequencies of the coherent noise terms have been identified from the power spectral density of $y_s(x)$. For each frequency, the variation of $s(x)$ with respect to (x) can be written

$$\begin{bmatrix}\dot{s}(x)\\ s(x)\end{bmatrix} = \Phi \begin{bmatrix}\dot{s}(o)\\ s(o)\end{bmatrix} \tag{47}$$

The values of $s(o)$ and $\dot{s}(o)$ determine the amplitude and phase of the signal $s(x)$. If we consider the vector on the right side of (47) to be a constant parameter vector h_o, then we can estimate h_o to minimize a least squares error where the i^{TH} error is

$$e_i = y_s(x_i) - [01]\, \Phi_i \hat{h}_o$$

The result of determining $\hat{h}_o$ is a function $\hat{s}(x)$ which, when subtracted from $y_s(x)$, results in an estimate of $y(x)$ which is optimal in a least squares sense.

10.0 Conclusions

It has been demonstrated that the conventional weighted least squares estimate is biased when applied to the structural parameter estimation problem. The three methods of bias removal, in order of effectiveness, are

1. LITWELS
2. IV
3. Subtraction

The LITWELS method and the subtraction method require the use of noise covariance matrices. When these matrices must be estimated, the LITWELS technique is somewhat degraded whereas the subtraction method may actually be improved at the expense of greater complexity. The IV method does not use the covariance matrices for bias removal, but the IV must be generated to be uncorrelated with the noise terms and correlated with X_e. If minimum variance is desired for the IV method, the noise covariance matrices must be known or derived.

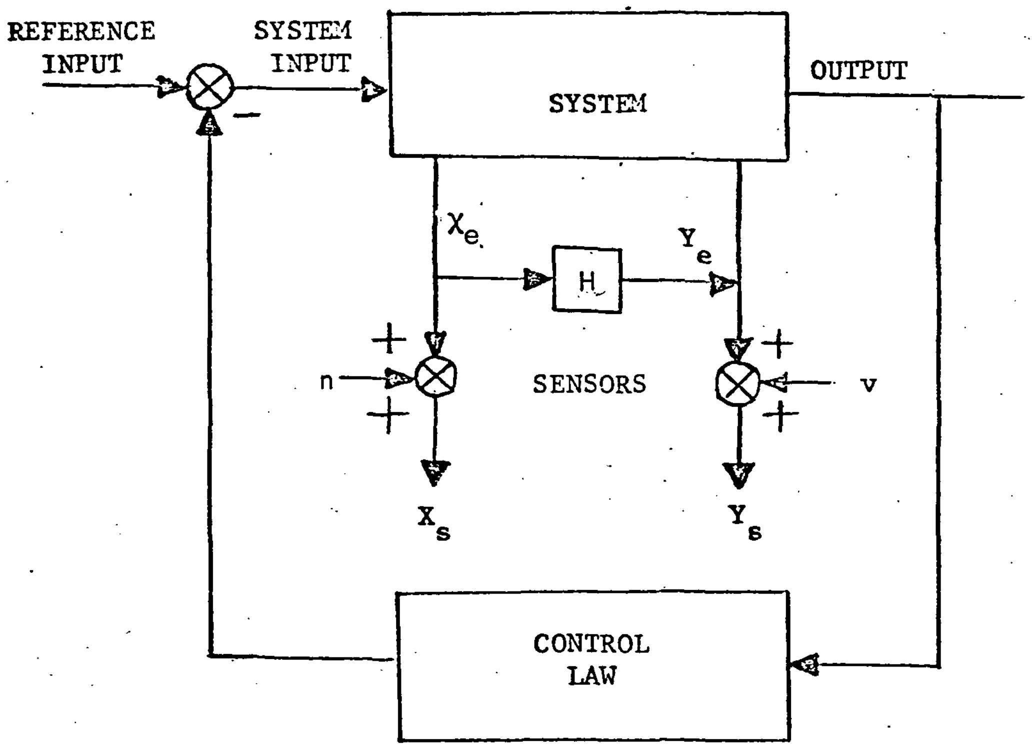

Fig. 1. Basic System Block

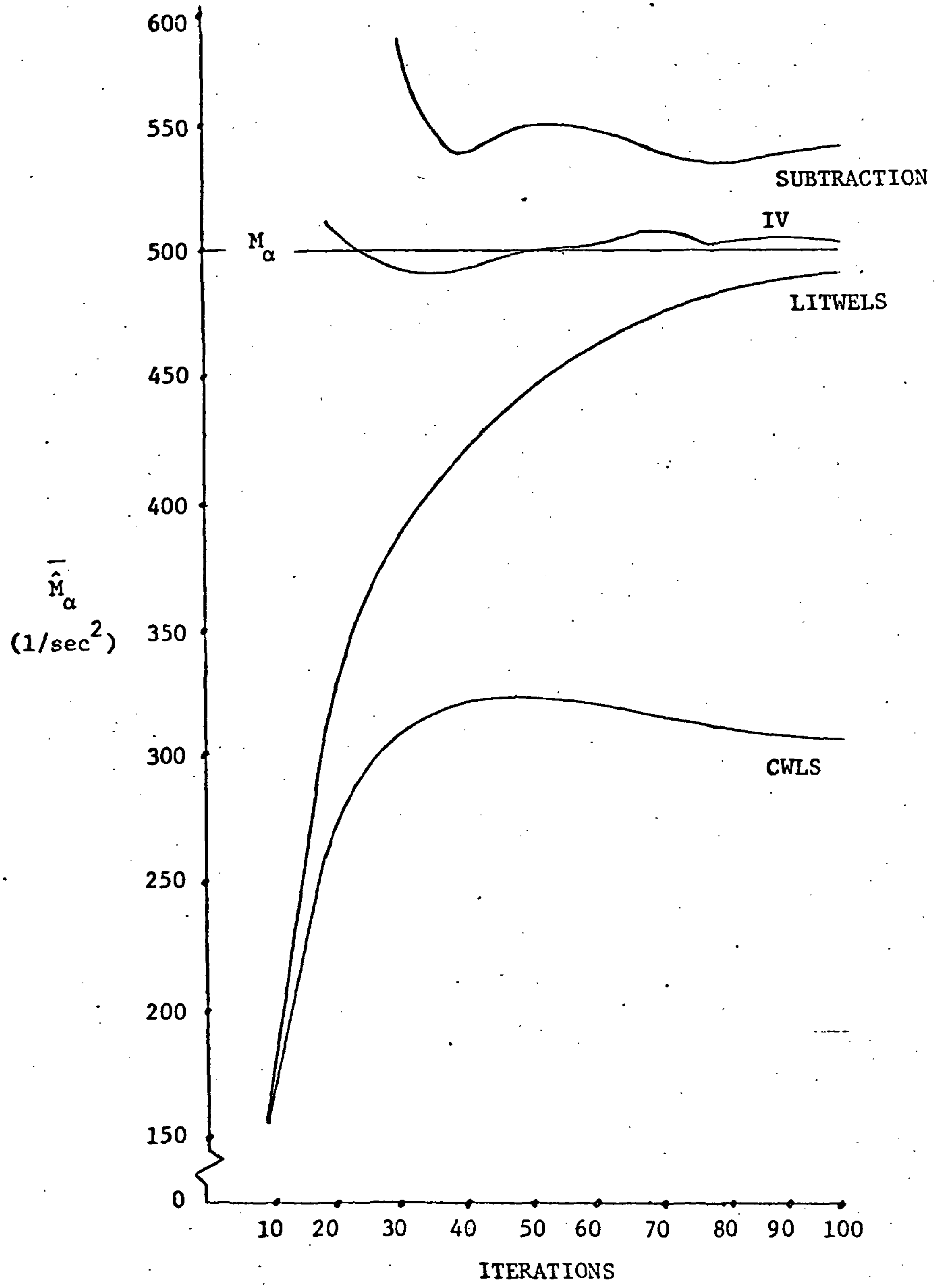

Fig. 2. Sample Mean of $\hat{M}_\alpha$ (Statistics Known)

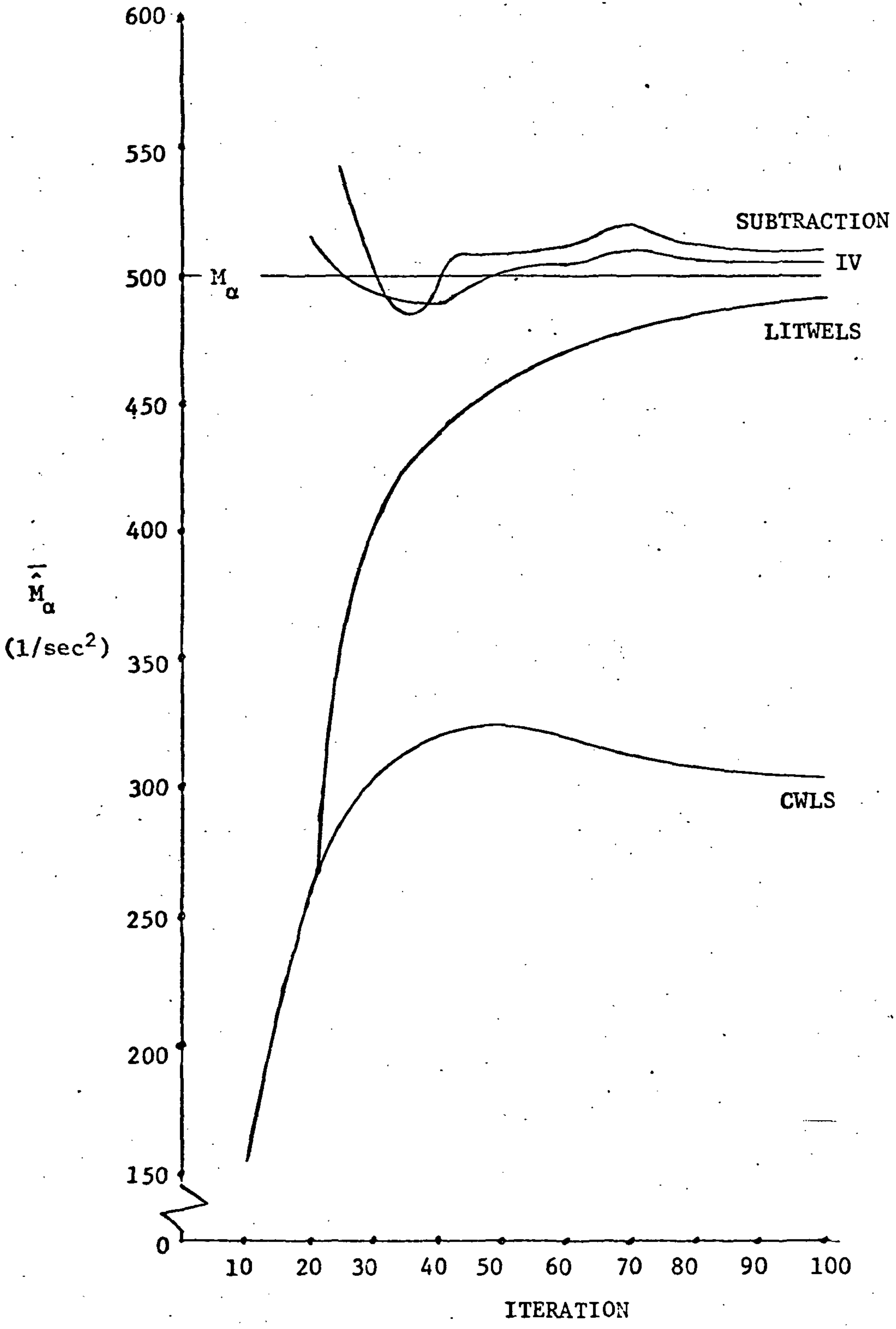

Fig. 3. Sample Mean of $\hat{M}_\alpha$ (Statistics Known)

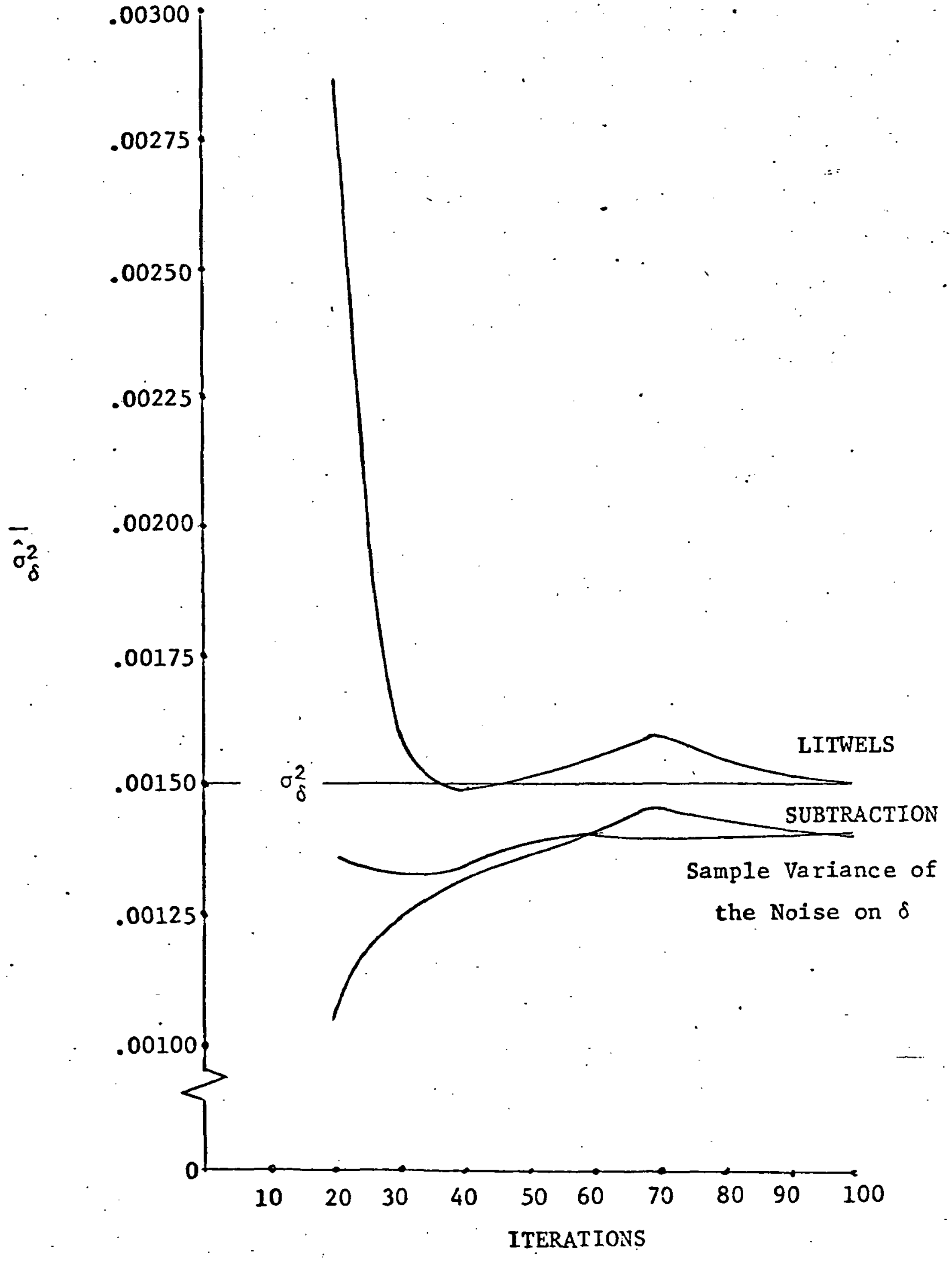

Fig. 4. Sample Mean of $\sigma_{\partial}{}^{2}$

REFERENCES

Problem III - Connection Between Kalman Filtering and Least Squares Estimation

1. Fagin, S. L. "Recursive Linear Regression Theory, Optimal Filter Theory, and Error Analysis of Optimal Systems." Sperry Inertial Systems Division, 1964.

2. Sage, A. P. "Optimum Systems Control." Prentice-Hall, Inc., 1968, Chapter 10.

3. Sorenson, H. W. "Least-Squares Estimation: From Gauss to Kalman." IEEE Spectrum, July 1970, pp. 63-68.

Problem III - Recursive Optimal Estimation

4. Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems." ASME Transactions, Journal of Basic Engineering, Vol. 82D, March 1960, pp. 35-45.

5. Kalman, R. E. and Bucy, R. C. "New Results in Linear Filtering and Prediction Theory." ASME Transactions, Journal of Basic Engineering, Vol. 83D, March 1961, pp. 95-108.

6. Nohi, N. E. "Estimation Theory and Applications." John Wiley and Sons, 1969.

7. Young, P. C. "Applying Parameter Estimation to Dynamic Systems - Part I." Control Engineering, October 1969, pp. 119-125.

Problem IV - Achieving Minimum Variance Estimators

8. Anderson, R. L. and Bankroft, T. A. "Statistical Theory in Research." McGraw-Hill Book Company, 1952.

9. Scheffe, H. "The Analysis of Variance." John Wiley and Sons, Inc., 1959, pp. 8-21.

Problem IV - Minimum Variance Estimators and Least Squares Estimators

10. Golub, G. H. "Comparison of the Variance of Minimum Variance and Weighted Least Squares Regression Coefficients." Annals of Mathematical Statistics, Vol. 34 (1963), pp. 984-991.

11. Magness, T. A. and McGuire, J. B. "Comparison of Least Squares and Minimum Variance Estimates of Regression Parameters." Annals of Mathematical Statistics, Vol. 33 (1962), pp. 462-470.

12. Watson, G. S. "Linear Least Squares Regression." Annals of Mathematical Statistics, Vol. 38 (1967), pp. 1679-1699.

13. Zyskind, G. "On Canonical Forms, Non-Negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models." Annals of Mathematical Statistics, Vol. 38 (1967) pp. 1092-1109.

14. Goodman, A. F. "Extended Iterative Weighted Least Squares Estimation of a Linear Model in the Presence of Complications." McDonnell Douglas Astronautics Company, Douglas Paper 5205, January 1969.

15. Rosenberg, A. E. and Shen, D. W. C. "Regression Analysis and Its Application to the System Identification Problem." Proceedings, Joint Automatic Controls Conference, 1963, pp. 446-451.

Problem II - The Bias Problem

16. Levin, M. J. "Estimation of System Pulse Transfer Function in the Presence of Noise." Proceedings, Joint Automatic Controls Conference, 1963, pp. 452-458.

17. Young, P. C. "Regression Analysis and Process Parameter Estimation, a Cautionary Message." Simulation, Vol. 10, No. 3, March 1968, pp. 125-128.

Problem II - The Instrumental Variable Method

18. Andeen, R. E. and Shipley, P. P. "Digital Adaptive Flight Control System for Aerospace Vehicles." AIAA Journal, Vol. 1, No. 5, May 1963, pp. 1105-1109.

19. Kendall, M. G. and Stuart, A. "The Advanced Theory of Statistics - Volume 2." C. Griffen and Co., Ltd., 1961, pp. 397-408.

20. Young, P. C. "An Instrumental Variable Method for Real-Time Identification of a Noisy Process." Automatica, Vol. 6, 1970, pp. 271-289.

Missile Flight Dynamics

21. Douglas Aircraft Company. "Upstage Flight Control." DAC-58865. December, 1967.

[illegible]

22. Waymeyer, W. C. and Young, T. H. "Coupling in Cruciform Missile Control Systems." AIEE Winter Group Meeting, New York City, New York, January 29 to February 2, 1962.

Adaptive Flight Control

23. Stefani, R. T. "Design and Simulation of a High Performance, Digital, Adaptive, Normal Acceleration Control System Using Modern Parameter Estimation Techniques." Douglas Aircraft Company Report DAC-60637. May 1967.

24. Stefani, R. T. "Comparison of the Dynamic Behavior of Single-Axis Controllers, Analog and Digital." Douglas Aircraft Company Report DAC-62380. November 1968.

Applications

25. Sorenson, H. W. "Least-Squares Estimation: From Gauss to Kalman." IEEE Spectrum, July 1970. pp. 63-68.

26. Kalman, R. E. "Design of a Self Optimizing Control System." Transactions, ASME, Vol. 80, February 1958. pp. 468-478.

27. "With Mechanical Programmer, New VSTOL has Simpler Controls." Product Engineering. October 20, 1969. p. 16.

28. "Piggyback Design is Strong Contender for Shuttle Contract." Product Engineering. January 19, 1970. pp. 14-15.

29. "Space Vehicle Will be Designed for the Nuclear Rocket Engine." Product Engineering. May 11, 1970. p. 10.

30. Rasool, I. S. "Evolution of the Atmosphere of Earth and Venus." Proceedings of the Symposium at Goddard Space Flight Center.

31. "Adaptive Controls Bring Soft Touch to Automation." Product Engineering, June 30, 1969. pp. 43-44.

32. "Brain Controls Use of Articial Arm." Product Engineering. October 7, 1968. p. 23.

33. "Novel Artificial Arm Gives Wearer Six Degrees of Freedom." Product Engineering. October 20, 1969. p. 15.

34. Montgomery, D. R. "Optics of the Mariner Imaging Instrument." Applied Optics. February 1970. pp. 277-287.

35. Billingsley, F. C. "Application of Digital Image Processing." Applied Optics. February 1970. pp. 289-300.

36. Papoulis, A. "Systems and Transforms with Applications in Optics." McGraw-Hill Book Company. 1968.

37. Cummins, H. and Midlo, C. "Finger Prints, Palms, and Soles." Dover Publications. 1961.

ON SEQUENTIAL SEARCH FOR THE MAXIMUM OF AN UNKNOWN FUNCTION

S. Yakowitz

1. INTRODUCTION

Many problems arising in engineering and operations research contexts have the following structure: The decision maker is provided with a class F of functions, whose common domain, X, is specified. Some mechanism selects a function f from F. The decision maker is not informed of this choice. He would like somehow to find a point $x^* \epsilon X$ at which f assumes its maximum value (denoted by $||f||$). Toward this end, the decision maker may sequentially and without constraint select elements $x_1, x_2, \ldots$ from X. Upon choosing x_n, he is informed of the value $f(x_n)$. Thus he may come to learn certain features of f. Any (perhaps randomized) strategy for choosing x_n on the basis of the sequence of pairs $\{(x_j, f(x_j))\}_{j=1}^{n-1}$ will be termed a search procedure. The problem of finding a search procedure S under which, for all $f \epsilon F$, $\{f(x_n)\}$ converges to $||f||$, in some specified sense, has generated a lively body of research papers, some of which will be referenced and described in the present paper.

As an example of the sort of engineering question giving rise to a search problem, suppose that an airplane is to fly with a fixed velocity. Its fuel efficiency will then be a function of the carburation setting. If x is the relative mixture of fuel and air and f(x) the associated fuel consumption required to maintain the aircraft's velocity, then the framework for a search problem is present. For this problem, X may be taken to be the unit interval and F, perhaps, may be considered to be the set of continuous functions on the unit interval.

Under certain restrictions on F and X, effective search procedures have been revealed. The most publicized of these is the "gradient method" which, in its simplest form, determines x_{j+1} from x_j by estimating the gradient ∇f of f at x_j (by difference approximations derived from local samples) and then setting $x_{j+1} = x_j + \lambda \nabla f(x_j)$. λ is a scalar chosen from heuristic considerations and may vary as the process evolves. If the functions in F are concave or at least unimodal and X is bounded and sufficiently regular, the gradient method can perhaps provide a Cauchy sequence $\{f(x_j)\}$ converging to $||f||$. Hadley's book Nonlinear and Dynamic Programming [1] devotes a nicely written chapter to the gradient method and its variations. The review paper by Spang [2] has an extensive bibliography on the gradient method, more recent techniques of which are described in the book by Osborn and Kowilak [3].

J. Kiefer [4,5] has published interesting analyses for the case that X is a bounded interval in the real line. In particular, under the search procedure he proposes, in n trials (the number n must be specified in advance) the point x^* at which $f(x^*) = ||f||$ can be located within a distance of $1/L_n$, L_n being the nth Fibonacci number, when F is the set of unimodal functions on $[0,1]$. Further, the search procedure is minimax in the sense that no non-randomized strategies can improve on this operating point error uniformly in F. Bellman and Dreyfus [6] devote a chapter to this optimization approach. To this writer's knowledge, an analgous search which also possesses the minimax property has yet to be revealed for multi-dimensional X.

An intriguing search model (which is slightly closer to the path to be followed here in that probabilistic ideas are prominent and multi-modal functions are included in F) was proposed by H. Kushner [7,8] who supposed f to be a sample function from a Brownian motion process on a bounded linear interval, $\mathcal{X}$. An advantage to this viewpoint is that, in addition to including multi-modal functions, ideas from Wiener prediction theory can be brought to bear on the problem of designing an optimal search procedure. Kushner points out that numerical evaluation of the optimal procedure is computationally prohibitive, but suggests (without proof) a search procedure under which $\lim_{n\to\infty} 1/n \Sigma_{i=1}^{n} f(x_i) = \|f\|$, almost surely.

The research reported in this paper follows an approach sketched by S. Brooks [9]. Presumably, Brooks took X to be a bounded subset of a Euclidean space, and the loss associated with the function $f\varepsilon\mathcal{F}$ and operating point $x\varepsilon\mathcal{X}$ to be

$L(x,f)$ = relative (with respect to $\mathcal{X}$) volume of points x' such that $f(x') > f(x)$.

Then, given any positive numbers and, a smallest number N is readily calculated such that if $X_1, X_2 \ldots X_N$ are selected uniformly from $\mathcal{X}$. Then for any real-valued function f,

$P[\max_{1\leq i\leq n} L(X_i,f) > c] < d$, for $n \geq N$.

Brooks, as well as Kushner, consider the possibility that the measurements $\{f(X_i)\}$ may be corrupted by additive noise. These considerations will be detailed, along with a brief review of "stochastic approximation" in a later section (Section 4) of this paper.

Let us loosely summarize the results of our investigation. (X,A) will be a measurable space, and M, the set of measurable functions on X. P is a probability function on (X,A). Examples show that no search procedure achieves $f(X_i)\to\|f\|$, even over the continuous functions on the unit interval or functions f on a countable $\mathcal{X}$. However, a search is presented such that for all $f \varepsilon M$, $f(X_n) \to f$ in P-probability. Also, we reveal a search which achieves P-almost-sure convergence to f of the terms $1/n \sum_{i=1}^{n} f(X_i)$.

The section closes with a description of various important subsets of M for which there are searches achieving Cauchy convergence of $\{f(X_i)\}$ to $||f.||$

Generalizing the idea of Brooks mentioned above, Section 3 proposes, as the loss associated with operating point $x \varepsilon X$ and criterion $f \varepsilon M$, the function $L(x,f) = P[\{y:f(y) > f(x)\}]$. The motivation is that we are able to derive upper bounds on the number of search iterations needed to achieve some given level of performance. Specifically given positive numbers c and d, we compute searches S_1 and S_2 and numbers N_1 and N_2 such that under S_1, if $n > N$

$$P[L(X_n,f) > c] < d;$$

under S_2,

$$P[\sup_{n>N_2} 1/n \sum_{i=1}^{n} f(X_i,f) > c] < d$$

Section 4 generalizes the search problem previously discussed by allowing that the observations of $f(x)$ may be corrupted by measurement noise. In the first theorem of the section, it is shown that if the measurement noise is additive and identically and independently (of x and f(x) as well as previous samples) distributed, then under our search procedure (which is independent of the noise distribution)

$$f(X_i) \to ||f||, \text{ in P-probability,}$$

and

$$L(X_i,f) \to 0$$

in P-probability. Also, for positive ε, c, and d, a procedure is revealed for finding the sample size N such that, under the search described, if $n>N$

$$P[\ y:f(y) > f(x_n) + \varepsilon \quad > c]<d.$$

This last result does require that the noise distribution be known. In Theorem 4.6, the noise is allowed to depend on x and f(x), but various assumptions are made about its mean, median, and variance.

2. ON THE EXISTENCE OF CONVERGENT SEARCHES.

We first introduce the notation and terminology to be used in the sequel. Let (X,A) be a measurable space and M the set of real-valued measurable functions or X. We shall always assume that each singleton set is in A. Let G be a subset of M and let $\|f\| = \sup_{x\epsilon X} f(x)$ for $f\epsilon M$ ($\|f\| = +\infty$ is possible). (We note that $\|f\|$ is not a true norm as it may be negative, for example.) A *deterministic search procedure* is a collection of measurable mappings $(m_k; k = 0, 1, 2, \ldots$ of $X^k X R^k$ into X (where R is the real line). Given a deterministic search procedure, for $f\epsilon G$ define inductively $x(0,f)=m_0$ and $x(k+1,f) = m_k\ (x(0,f), \ldots, x(k,f), f(x(0,f)) \ldots f(x(k,f))$. We say that G has a *deterministic search* if a deterministic search exists such that $\lim_{n\to\infty} f(x(n,f)) = \|f\|$ for all $f\epsilon G$. Of course, the intuition behind this definition is that $x(n,f)$ is the next point at which we observe the value of f after having observed $f(x(j,f))$, $j=1,2,\ldots, n-1$.

A *random search procedure* consists of a mapping $m_k(B;x_1, \ldots, x_k, y_1,\ldots,y_k)$ defined for $B\epsilon A$ and $x_i\epsilon X$ and y_i R, k=0, 1, 2, ... Further for fixed, $x_1, \ldots, x_k, Y_1, \ldots, y_k$, $m_k(.;x_1, \ldots, y_k)$ is a probability measure on (X,A) and for fixed $B\epsilon A$, $m_k(B\ ;\ .\ )$ is a measurable function on $X^k X R^k$.

We interpret $M_k(.;\ x_1, \ldots, x_k, y_1, \ldots, y_k)$ as the conditional probability distribution of X_{k+1} if we observe $x_1, \ldots, x_k$, $f(x_1) = y_1, \ldots, f(x_k) = y_k$. For each $f\epsilon G$ we may find a probability distribution on the sequence space

X^∞ by defining a consistent family of measures on X^k for each k. Let $P_1^f = m_0$ and inductively,

$$P_k^f\ (A \times B) = \iiint_B m_{k-1}\ (A;\ x_1, \ldots, x_{k-1}\ f(x_1), \ldots, f(x_{k-1})$$

$$P_{k-1}^f\ (dx_1\ dx_2\ , \ldots , dx_{k-1})$$

where $A \varepsilon \mathcal{A}$ and $B \varepsilon \mathcal{A}^{k-1}$. Let P_f be the resulting probability measure on X^∞.

We say that G has an <u>almost sure search</u> if there is a random search such that for all $f \varepsilon G$,

$$P_f(f(X_n) \to \|f\|) = 1,$$

where X_n is the identity function on the <u>n</u>th coordinate of X^∞. We say that G has a <u>search in probability</u> if for all $f \varepsilon G$, $\lim_n P_f(f(X_n) \ \varepsilon\ N(\|f\|)) = 1$ for each neighborhood $N(\|f\|)$ of $\|f\|$ (with the usual neighborhood system at infinity). If $\|f\|$ is finite this is the same as requiring that $\lim_{n\to\infty} P_f(|f(x_n) - \|f\|| < \varepsilon) = 1$ for each $\varepsilon > 0$.

In most spaces if we consider $G = M$, then it is too much to hope for any sort of convergence since a function may be large at a "small" set of points. To get around this trouble it is convenient to allow the function to be arbitrarily defined on a small set. Let P be a probability measure on $(X, \mathcal{A})$ and for each $f \ \varepsilon\ M$ let $\|f\|_P$ be the P-essential least upper bound of f. The measure P may take into account <u>a priori</u> knowledge of which x values are important, but we will not elaborate this point.

We say that G has a <u>P-almost sure search</u> if there is a random search such that for all $f \ \varepsilon\ G$,

$$P_f(\liminf f(X_n) \geq \|f\|_P) = 1.$$

We say that G has a P-search in probability if there is a random search such that $\lim_{n\to\infty} P_f\,(f(X_n) \in (||f||_p - \varepsilon\ , +\infty)) = 1$ for each $\varepsilon>0$. (The obvious modification holds if $||f||_p = +\infty$.)

We first consider two examples to show the trouble one may have finding search procedures.

Example 2.1: Let X be countable with each single point set in A. Let G consist of functions taking rational values on X. Let P put positive measure on each one point set. Then G does not have P-almost sure search.

Proof: Let f be the indicator function of $\{x\}$ where $P(\{x\}) > 0$. Find N such that $P_f(X_n = x\ , n \geq N) > 0$ and then find $x_1, . . , x_{N-1}$ such that $P_f(\{(x_1,.,x_{N-1},x,x,..)\}) > 0$. Consider $g(z) = 1$ if $z = x$, $z = 2$ if $z = y$ where $P(\{y\}) > 0$ and y $\{x_1, . . , x_{n-1}\}$, and $g(z) = 0$ if $z \neq x$ or y. Then

$P_g(g(X_n) \to ||g||_p) < 1.$

Example 2.2: Let $X = [0,1]$, A= The Borel field of $[0,1]$, G = continuous functions and P be Lebesgue measure. Then G does not have a P-almost sure search.

Proof: Let f have a unique maximum at 1. If $f(X_n) \to ||f||$ almost surely then there is some interval $I \subseteq [0,1/2]$ such that $P_f(X_n \not\in I, n = 1, 2, ...) > 0$. Consider any continuous function g which agrees with f outside of I and takes its maximum in I. It is easy to see that $g(X_n) \to ||f|| < ||g||$ with positive probability.

Note that in the two examples $||f||_p = ||f||$ for all $f \in G$. Further, since each single point set is in A any deterministic search procedure is also a random search procedure. Thus, in the two examples G does not have a deterministic search.

There are at least two ways of getting around this problem: 1) Consider smaller classes of functions, (e.g. unimodal [5]. 2) Use different criteria for convergence.

We now see that a P-search in probability is possible even when $G = M$.

<u>Theorem 2.3</u>: Let $G = M$. Then for each P, G has a P-search in probability.

<u>Proof</u>: Let $X_1, X_2, \ldots$ be independent identically-distributed random mappings each with distribution P. Then for each f

$$P_f\left(\max_{1\leq i\leq n} f(X_i) \to ||f||_P\right) = 1. \qquad (1)$$

The proof is completed by using the following lemma.

<u>Lemma 2.4</u>: Suppose that G has a random search such that (1) holds. Then G has a P-search in probability.

<u>Proof</u>: We sketch the proof. Let Y_i be Poisson random variables with parameters $\lambda_i \to +\infty$ which are mutually independent of each other and independent of $X_1, X_2, \ldots$. For each n, let $X_n^* =$ value of $X_1, \ldots, X_n$ which maximizes $f(X_i)$. Consider the random sequence

$$\underbrace{X_1, X_1^*, \ldots, X_1^*,}_{Y_1\text{ - times}}\ \underbrace{X_2, X_2^*, \ldots, X_2^*, \ldots,}_{Y_2\text{ - times}}$$

We can find "new" m_k which lead to the same distribution as the random sequence just given. It is easy to verify that this random search works to give convergence in probability.

The same method of proof yields:

Lemma 2.5: Let each $f \varepsilon G$ have P-ess inf $f(x) > -\infty$ and (1) hold; then G has a random search procedure such that

$$P_f \left(\frac{1}{n} \sum_{i=1}^{n} f(X_i) \to \|f\|_P\right) = 1$$

for all $f \varepsilon G$.

In reference to Theorem 2.3 it is the opinion of the authors that in practice, convergence in probability is as useful as convergence almost surely. In either case one would like some information on the rate of convergence (a point we return to later).

Let us turn to the other approach of finding searches by restricting the class G. The following results are all easy.

Lemma 2.6: Let X be an arbitrary topological space. If for every $\varepsilon > 0$ there is a finite collection $E_1, \ldots, E_n$ of sets with union X and points S_i in E_i, i=1, 2, .. , n such that

$$\sup_{f \varepsilon G} \quad \sup_{S \varepsilon E_i} \quad |f(S_i) - f(S)| < \varepsilon$$

then G has a deterministic search.

Proof: Let $\varepsilon_i = 1/2^i$, i=1, 2, ... Find $E_1(1), \ldots, E_{n(1)}{}^{(1)}$ sets associated with ε_1. Let $X_1 = x_1 \ \varepsilon E_1(1), \ldots, X_{n(1)} = x_{n(1)} \ \varepsilon E_{n(1)}(1)$. Let $f(X_1^*) = \max_{1 \leq j \leq n(1)} f(X_j)$. If $f(X_1^*) - f(X_j) \geq 2\varepsilon_j$ delete E_j from the space X.

all over again with the new space and ε_2, etc.

Corollary 2.7: Let X be compact, metric and A the Borel sigma-field. Then any equicontinuous family of functions G has a deterministic search.

Proof: As X is compact, G is uniformly equicontinuous.

Corollary 2.8: Let X be compact and metric and A the associated Borel sigma-field. Then any compact subset (with respect to the sup-norm metric) of the continuous functions on X has a deterministic search.

Proof: Ascoli-Arzela theorem and Corollary 2.7.

Corollary 2.9: Let X be compact and metric and G uniformly satisfy a Lipschitz condition. Then G has a deterministic search.

Proof: By Corollary 2.7.

Corollary 2.10: Let X be a compact, differentiable manifold, A the generated sigma-field and all f in G have uniformly bounded derivatives. Then G has a deterministic search.

Proof: By Corollary 2.7 since the family is equicontinuous.

Corollary 2.11: Let X be a compact metric space, A the Borel field of X and let $C(X)$ be the continuous functions on X with the sup-norm. Let μ be a probability measure on $C(X)$ with its Borel sigma-field. Then for each $\varepsilon > 0$ a deterministic search may be found such that

$$\mu(f : f(X_n) \to ||f||) > 1 - \varepsilon.$$

Proof: First note that we may think of f as a sample path from a stochastic process with domain X. From the conditions of X it follows that $C(X)$ is a complete separable metric space ([10, pp 94,103]), and hence μ is a tight measure [11). Thus, we may find a compact set K such that $\mu(K) > 1 - \varepsilon$. Use Corollary 2.8 on K.

Observe that if $X = [0,1]$, the sigma-field of this process is the same field that is generated by the usual "product-field" construction (a proof of this statement is in Parasarathy [12] p. 212). In particular, Corollary 2.11 is related to a study by Kushner [7] which proposes a search for finding the maximum of Brownian motion sample functions.

The problem of characterizing subsets G that have a deterministic search is quite interesting, but the authors have not been able to make much progress. The problem appears to lie in the domain of mathematical logic.

Note that under the conditions of Corollary 2.11 we may find a countable dense set of points in X. Let P be a measure putting positive mass on each point of the set. Then for continuous f, $||f|| = ||f||_P$ and Theorem 2.3 and Lemma 2.5 hold for the stochastic process.

3. RATES OF CONVERGENCE

In applications of sequential search procedures it is highly desirable that there be some way of assessing what can be done in a finite number of iterations. For example, one would be interested in knowing, if possible, how fast $f(X_n) \to ||f||$. In this section we consider questions of this sort.

To see the difficulties involved we consider an example in random search.

Example 3.1: As a criteria of the amount of convergence one might consider $||f|| - f(X_n)$ or $(||f|| - f(X_n))/||f||$. More generally, we will use $g(||f||, f(X_n))$ where $g(x,y)$ is a function satisfying:

1) for fixed x, $g(x,y)$ is strictly decreasing as y approaches x from below.

2. $g(x,x) = 0$

3. For fixed y, $g(x,y)$ is strictly increasing as x increases (where $x \geq y$) with a limit ≥ 1 for all y as $x \to \infty$.

To get a grasp on the rate of convergence one might hope to find a random search such that for each $c > 0$ and $0 > d > 1$, there is a number $N(c,d)$ such that for $f \varepsilon G$ and for $n \geq N(c,d)$

$$P_f(g(\|f\|, f(X_n)) > c) < d \text{ .}$$

We now show that this cannot be done if $X = [0,1]$ and G is the set of continuous functions. Let $c \leq 1/2$, $0 < d < 1$, and n be any fixed integer and some random search procedure also be fixed. Pick any $f \varepsilon G$ and let I be an interval such that

$$P_f(I \cap X_1, \ldots, X_n = \emptyset) > d.$$

Let $h \varepsilon G$ agree with f on the complement of I and $g\|h\|, \|f\|) > 1/2$. Then we have

$$P_g(g(\|h\|, h(X_n)) > c) \geq P_g(g\|h\|, h(X_n)) > 1/2)$$

$$\geq P_g(h(X_i) = f(X_i), i=1, 2, . \, , n)$$

$$\geq P_g(X_1, . \, . \, , X_n \cap I = \emptyset) = P_f(X_1, . \, . \, ,X_n \cap I = \emptyset)$$

$$> d$$

ending the example.

A great weakness in the theory of search procedures is the fact that for G the class of continuous functions on $[0,1]$, under no search procedure can bounds on the rate of convergence of $f(X_n)$ to $\|f\|$ or $\Sigma_{i=1}\ f(X_i)/n$ to $\|f\|$ be established which are uniform on G. The practical consequence of this weakness is that the experimenter cannot estimate the level of performance attainable in a finite number of search iterations. One approach to overcoming these difficulties is to redefine the search problem by proposing a different (but, hopefully, not unreasonable) criterion of goodness.

We do this by following some of the ideas implicit in Brooks [9.]. Associated with each operating point $x \varepsilon X$ and $f \varepsilon G$ is the set $\alpha(x,f) = \{y:f(y) > f(x)$, which is here called the domain of improvement (of f over f(x)).

As in section II, let a probability measure P be given on $\mathfrak{A}(X,A)$. As a loss function we propose the P measure of $\alpha(x,f)$. That is,

$$L(x,f) = P\ (\alpha(x,f)).$$

Strictly speaking, L should also contain P as a variable, but since P will be fixed we shall omit this notation.

Thus, $L(x,f)$ is the probability that a person choosing a point Y at random in X with distribution P will find that $f(Y) > f(x)$. If X is a set of finite volume in R^k and P is proportional to volume (i.e., P is proportional to Lebesque measure) then $L(x,f)$ is the fraction of the volume on which f exceeds $f(x)$.

We find that for certain search procedures it is possible to obtain information on how close $L(X_n,f)$ is to zero. We will say that $X_1, X_2, \ldots,$ are chosen at random if $X_1, X_2, \ldots$ are independent, identically distributed X- valued random mappings with distribution P. Let $f \in M$ be fixed and for each n define n^* by $1 \leq n^* \leq n$ and

$$f(X_{n^*}) = \max_{1 \leq i \leq n} f(X_i)\ .$$

<u>Proposition 3.2</u>: Let $X_1, X_2, \ldots$ be chosen at random and $0 < a < 1$; then for each integer n and $f \in M$,

$$P_f(L(X_{n^*},f) > a) \leq (1-a)^n.$$

<u>Proof</u>: Let $t' = \sup \{t\colon P(\{x{:}f(x) > t\}) > a\}$ then $P(\{x{:}f(x) \geq t'\}) \leq a$. Thus, $P_f(L(X_{n^*},f) > a) = P_f(f(X_i) < t',\ 1 \leq i \leq n) = \prod_{i=1}^{n} P(\{x{:}f(x) < t'\}) \leq (1-a)^n$. By considering any random variable $f(X)$ with a continuous distribution function we see that equality may hold.

With this criterion of accuracy, the rate of convergence is independent of the dimensionality of the space X. With other criteria of convergence this might not be true. This point is discussed further by Spang [9, page 362] who uses two concepts of convergence and appears to doubt Brooks' [9] comment that the rate of convergence is independent of dimensionality.

Further information on the rate of convergence is contained in the next theorem. In what follows, M_n is defined to be the random variable $L(X_{n*},f)$ defined in Proposition 3.2, and X is distributed randomly.

<u>Theorem 3.3</u>: Let f(X) have a distribution function F such that for some $\varepsilon>0$, $F(x) \geq 1 - \varepsilon$ implies F is continuous at x. Then nM_n converges weakly to be exponential distribution with parameter 1.

<u>Proof</u>: Let a > 0; then for large n,

$$P_f(nM_n \leq a) = P_f\ (M_n \leq a/n)$$

$$= 1 - P_f(M_n > a/n) = 1 - \Pi^n_{i=1}\ P_f(L(X_i,f) > a/n)$$

$$= 1 - \Pi^n_{i=1}\ P\ (\{x: f(x) > t_n\}) = 1-(1-a/n)^n$$

where $t_n = \sup\{t: P(\ x:\ f(x) > t\}) > a/n$. This approaches $1 - e^{-a}$ as $n \to \infty$ completing the proof. For a > 0, by Taylor's theorem with remainder on the logarithm of $e^{-x}/(1 - x/n)^n$, we see that (large n):

$$\exp(-a^2/2n) < e^{-a}/(1-P_f(nM_n \geq a)) < \exp(-a^2/2n + a^3/6n^2)$$

In the same vein as Lemmas 2.4 and 2.5, it is shown that we can use Proposition 3.2 to get searches which converge at a known rate.

<u>Theorem 3.4</u>: One may compute a search procedure S_1 under which, for any positive numbers c and d, a number N(c,d) may be found for which

$$P[\sup_{n>N(c,d)} 1/n \sum_{i=1}^{n} L(X_i,f) > c]<d$$

for every $f\varepsilon M$.

<u>Proof</u>: Let $\{n(i)\}_{i=1}^{\infty}$ be a sequence of numbers such that $n(1) = 1$ and $1/n(i)$ converges to 0 monotonically (e.g. 2^{i-1}). By Proposition 3.2, we may compute a number N' such that

$$(c/2)\ [n(N')-N'')/n(N'')] +1[n(N')+N'')/n(N'')] < c.$$?

Search procedure S_1 requires that *X* be sampled independently with distribution P at times $t=n(j)$ $(j=1,2,\ldots)$, and for $t\neq n(j), x_t$ is chosen to be the best value in the sequence $\{X_{n(j)}\}$ sampled thus far: $f(x_t) = \max\{f(X_v) : v \leq t\}$. Thus evidently $f(x_t)$, $t \notin \{n(j)\}$ is monotonically increasing in t. Observe that from the choice of N' and the definition of S_1,

$$P[L(X_{n(N')},f) > c/2] < d.$$

Let Q be the event (with reference to the process determined by S_1 on f) that $L(X_{n(N')},f) \leq c/2$. If Q occurs, then by the choice of N" (and observation that $L(x,f) \leq 1$, always)

$$\sup_{n>N''} \sum_{i=1}^{n} 1/n\ L(X_i,f) \leq c.$$

In summary,

$$P[\sup_{n>N''} \sum_{i=1}^{n} 1/n\ [L(X_i,f) > c] \leq P\ [Q^c] < d,$$

and consequently the theorem is proved, with the understanding that N' suffices for N(c,d).

Theorem 3.5: One may compute a search procedure S_2, under which, for any positive numbers c and d, a number N(c,d) may be found for which

$$P[L(X_n,f) > c] < d$$

for all $n > N(c,d)$ and all $f \varepsilon M$.

Proof: Let $\{n(j)\}$ be a sparse sequence as in the proof of Theorem 3.4. From this we construct a random sequence $\{N(j)\}$ where $N(j)$ has the sample space $\{n(j), n(j)+1, n(j)+2, \ldots, n(j+1)-1\}$ and is chosen by the randomization which assigns equal probability to each element of this sample space. S_2 is the search procedure which samples X independently and uniformly at times in $\{N(j)\}$. At other times, x_t is chosen to be the best operating point thus far sampled. The condition imposed on $\{n(j)\}$ that $1/n(j)$ converge monotonically to 0 as j tends to infinity ensures us that a number N' can be found such that

$$P[N(j)=n] < d/2 \quad \text{for all } j > N', \text{ all integers } n.$$

From Theorem 4, a number N" may be found such that $P[M_{N''} > c] < d/2$. If $k = \max\{N'+1, N''\}$ then for $n > n(k)$

$$P[L(X_n,f) > c] \leq P[M_{N''} > c] + P[n \varepsilon \{N(j)\}] < d$$

ending the proof.

Without going into detail it is clear that in results 2.6 through 2.10 one may find integers $N(\varepsilon)$ such that if $n > N$, $||f|| - f(X_n) < \varepsilon$ for all $f \varepsilon G$. A modification of Corollary 2.11 also holds). In order to find N one must know quite a bit about the structure of G. For example, in Corollary 2.8 *one must know* the compact set. In general, there is no one search which works for all compact sets. If one knows the compact set, not only may a convergent deterministic search be found but also a uniform bound on the time necessary for any degree of convergence.

In closing we note that if G consists of uniformly bounded measureable functions, then the central results of this section obtain under the loss function $L'(x,f) \overset{\Delta}{=} \int_B f(y)P(dy)$, where $B = \{y:f(y) \lesssim f(x)\}$

Proposition 3.6: Let g be a function in $\mathcal{M}$ having a finite expectation with respect to P. Assume for every $f\varepsilon G$, $f \leq g$, and that X_i is a random sequence. Then for every positive c and d, a number N may be computed such that for every $f\varepsilon G$, in the notation of Proposition 3.2.

$$P[L'(X_{N*},f) > c] < d.$$

Proof: By the assumption that the integral of g is finite, one can find a positive number k such that if $P[A] < k$,

$$\int_A g(x)\, P(dx) < c.$$

If N is such that $(1-k)^N < d$, from the proof of Proposition 3.2, we know that with probability greater than 1-d, $L(X_{N*},f) < k$. By the definition of L and the choice of k, this means that with probability greater than 1-d,

$$\int_A g(x)P(dx) < c, \qquad (A = \ x:f(x) > f(X_{N*}) \qquad (3.1)$$

Finally, as g majorizes f, (3.1) gives us (letting $A \equiv \ x:f(x) > f(X_{N*})$

$$L'(X_{N*},f) = \int_A f(x)P(dx) \leq \int_A g(x)P(dx) < c$$

From this proposition, the other major results of Section III follow with L' replacing L, with at most minor modifications of the proofs.

4. SEQUENTIAL SEARCH USING NOISY MEASUREMENTS

In this section we consider the problem of the earlier sections with the additional complication that errors of measurement are present. To be more specific, if X_n is the nth operating point, the decision-maker observes:

$$f(X_n) + Z_n(X_n, f(X_n)) \tag{4.1}$$

where $Z_n(X_n f(X_n))$ is a random variable conditionally independent of $X_1, \ldots, X_{n-1}$, $Z_1, \ldots, Z_{n-1}$ (given X_n and $f(X_n)$) whose distribution is conditional on the values of X_n and $f(X_n)$. We will assume that if $X_i = X_j$, then Z_i and Z_j have the same distribution.

Physically, $f(X_n) + Z_n$ may be regarded as arising from a noisy meter which measures $f(X_n)$, the noise being dependent upon the operating point X_n and $f(X_n)$ the value at X_n. "Noisy measurements" refer to observations of the form (4.1) (in contrast to $f(X_n)$ which is considered a "noiseless measurement").

The basic idea in the section is the standard one of replicating observations to minimize the effect of observational error (see. e.g. Brooks [9]). We consider several different cases. The first case is when the measurement error does not depend upon X_n or $f(X_n)$. The distribution F_z of the error is assumed unknown in the next theorem.

Lemma 4.1: In the noisy measurement case, let $Z_1, Z_2, \ldots$ be unknown independent identically distributed random variables independent of $(X_1, f(X_1))$ $(X_2, f(X_2))$, .. for each $f \in M$. One may compute a search procedure S_3 under which, with P-probability 1, as $n \to \infty$,

$$1/n \sum_{i=1}^{n} L(X_i, f) \to 0$$

$$\text{(thus } 1/n \sum_{i=1}^{n} f(X_i) \to ||f||_p \text{ if } f \text{ is P-bounded below)}$$

for each $f \in M$ such that $P[f(X) = ||f||] = 0$.

Remark: For piecewise continuous functions f, this last restriction is satisfied if f does not assume its maximum on a plateau.

Proof: The description of the search procedure S_3 uses the following notation: $\{u(n)\}$ is an observation of a sequence of independent values $\{U(n)\}$ P-distributed on $\mathcal{X}$. Class $F_{N,j}$ denotes the empiric distribution function constructed from the observations which, during the first N observations of the search, have been made at $u(j)$, $j = 1,2,\ldots$. (An empiric distribution function F_n constructed from any sequence $\{x_i\}_{i=1}^{n}$ of n real numbers is the cumulative distribution function determined by the expression

$nF_n(x)$ = number of elements x_j of $\{x_i\}_{i=1}^n$ such that $x_j \leq x$.

$F_{u(j)}$ is the cumulative distribution function (cdf) for the random variables $f(u(j)) + Z$; i.e.,

$$F_{u(j)}(z) = F_Z(z+f(u(j))), \text{ for every real } z.$$

More generally, F_x is the cdf of $f(x) + Z$. If $H(x)$ is any real function, let $||H||^* = \sup_{x\in\mathcal{X}} |H(x)|$. $\{K(v)\}$ is a sequence of integers such that if $n>K(v)$, then for any cdf F, and empiric distribution function F_n constructed from n independent observations distributed as F,

$P[||F-F_n||\ * \geq 1/v] < 2^{-v}/v.$

Massey [13] gives an algorithm capable of computing a minimum such number $K(v)$.

$\{M(v)\}$ is a sequence computed inductively by the following rule:

$$M(2) = 1.$$

$$M(v) = M(v-1)+A(v)+v\ K(v), \quad v > 2$$

where $A(v)$ is some positive integer such that

$$[M(v-1)+v\ K(v)+(v+1)\ K(v+1)]/A(v) < 1/v \tag{4.2}$$

Having described $\{K(v)\}$ and $\{M(v)\}$, we are in a position to reveal the search procedure S_3.

Step 1:

For each iteration v, v = 2,3,..., of these Steps 1-3 the points $\{x_n\}_{n=M(v)}^{M(v)+vK(v)}$ are chosen, at each n, from the set of points $\{u(j): j = 1,2,..,v\}$, so that each u(j) is sampled K(v) times. Therefore, by time N = M(v) + vK(v).

$$P[\|F_{N,j} - F_{u(j)}\|^* \leq 1/v,\ j=1,2,..,v] > 1 - 2^{-v}. \qquad (4.3)$$

Step 2:

At time N = M(v) + vK(v), a positive integer $v^* \leq v$ is selected such that for every real number z,

$$F_{N,v^*}(z) > F_{N,k}(z) - 2/v \text{ for } 1 \leq k \leq v. \qquad (4.4)$$

If no such v* can be selected, v* is chosen arbitrarily.

Step 3:

At times n, $M(v) + vK(v)<n<M(v+1)$, $X_n = u(v^*)$. At time M(v+1), repeat the process, with v increased by 1. Toward outlining a proof that S_3, as just described, possess the property asserted in the theorem, it is necessary to recognize that with probability 1, (4.4) will hold for all but finitely many v. For demonstration of this, let u(v') be any positive integer not greater than v such that

$$f(u(v')) = \max_{1 \leq j \leq v} f(u(j)).$$

Then for all z and all $i \leq v$,

$$F_{u(v')}(z) = F_Z(z+f(u(v'))) \geq F_{u(i)}(z) = F_Z(z+f(u(i))).$$

The event (which will be denoted by B(v)) that

$$\|F_{N,j} - F_{u(j)}\|^* \leq 1/v\ ,\ 1 \leq j \leq v \qquad (4.5)$$

implies, by the triangle inequality, that for $j \leq v$,

$$F_{N,v'}(z) > F_{N,j}(z) - 2/v, \text{ all real } z$$

and thus (4.4) holds with $v^* = v'$. Note that by construction of $\{K(v)\}$,

$$\sum_{v=2}^{\infty} P(B(v)^c) < \sum_{v=2}^{\infty} 2^{-v} < \infty .$$

and consequently, by the Borel-Cantelli lemma, $B(v)$ occurs for all but finitely many v, concluding our assertion that for all but finitely many v, v^* can be picked to satisfy (4.4). We will hereafter assume without comment that v^* always has the property (4.4). As our only concern is with limit theorems, this assumption will not lead us astray.

The completion of the proof that S_3 leads to the convergence of $1/n \sum_{i=1}^{n} L(x_i,f)$ to 0 is at hand. By the choice of $M(v)$ and $A(v)$, we have that at all time Q during the vth iteration of steps 1-3 ($v>2$) that

[Number of Observations x_i, $1 \leq i \leq Q$, taken at $(v-1)^*$ or v^*]$/Q>(v-1)/v$,

and thus for all $n > M(3)$,

$$1/n \sum_{i=1}^{n} L(x_i,f) < 1/v+((v-1)/v) \max \{L(u(v^*),f),\ L(u((v-1)^*),f)\} \qquad (4.6)$$

The proof is completed by showing that almost surely,

$$L(u(v^*),f) \to 0.$$

Let x' be any point in X such that $L(x',f)>0$. Then almost surely some $u(h)$ in an observation of $\{U(v)\}$ gives $f(u(h)) > f(x')$. If H is a number such that

$$6/H < \| F_{x'} - F_{u(h)} \| .$$

Then for all $v > \max\{H,h\}$, if $f(u(j)) \leq f(x')$,

$$\begin{aligned} F_{N,v^*}(z) &\geq F_{u(h)}(z) - 2/v > F_{u(j)}(z) + 6/H - 2/v \\ &> F_{N,j}(z) + 6/H-4/v > F_{N,j}(z) + 2/v, \text{ (all real } z\text{)}, \end{aligned}$$

which implies that j cannot be chosen to satisfy (4.4) for v*. From this we deduce that

$$\lim\sup L(u(v^*),f" \leq L(x',f). \qquad (4.7)$$

Let $\{w_n\}$ be a sequence (whose existence is implied by the hypothesis that $P[f(X) = ||f||_P] = 0$) such that $L(w_n,f) > 0$ and $L(w_n,f) \rightarrow 0$. Then (4.7) holds almost surely simultaneously for all the w_n(in place of x') and we conclude that with probability 1,

$$\underline{\lim}\, L(u(v^*),f) \leq \inf_n L(w_n,f) = 0$$

<u>Theorem 4.2</u>: Under search S_3' described below, Lemma 4.1 remains true in the absence of the hypothesis $P[f(X) = ||f||_P] = 0$.

<u>Proof</u>: S_3' differs from S_3 only in step 2, where for S_3' the restriction is made that v* be the <u>greatest</u> positive integer $\leq$ v such that for every real number z,

$$F_{N,v*}(z) > F_{N,k}(z) - 2/v, \quad 1 \leq k \leq v. \qquad (4.8)$$

Observe that S_3' is a version of S_3, and consequently it achieves convergence under the hypothesis of the preceding theorem.

In the absence of a sequence $\{w_n\}$ as described in the proof of the previous theorem, there is a number t' such that

$$P[f>t'] = 0 \quad \text{and} \quad P[f = t'] > 0. \qquad (4.9)$$

(The abbreviation $P[f > b]$ is used to denote the P-probability of the domain of improvement $\{x:f(x) > b\}$). We use the notation of the proof to the preceding theorem. Let h be an integer (surely there is one) such that $f(u(h)) = t'$. Then for $v > h$, under S_3', v becomes v* by virtue of one of the events A(v) or B(v) (in the sigma-field of the process determined by S_3 and f) occuring:

$A(v)$ $\qquad f(\mathcal{U}(v)) = t'.$

$B(v)$: $\qquad B(v) = B_1(v) \cap B_2(v).$

where $\quad B_1(v)$: $\qquad t' > f(u(v)) \geq t' - a(v)$

and $\quad B_2(v)$: $\qquad F_{N,v}$ satisfies (4.4)

Here $a(v) = \inf \{a: || F_{t'} - F_a || * \leq 2/v\}$, $||$ * being the sup norm. Note that $P[A(v) \cup B(v)] \geq P[A(v)] = P[f{=}t']$ which is positive and independent of v. Thus under S_3', during evolution of the process infinitely many different v are chosen as v*. Our proof consists of showing (below) that

$$\lim_v P[B(v) \quad A(v) \cup B(v)] = 0 \tag{4.10}$$

Note that A(v) and $B_1(v)$ are independent of $\{\mathcal{U}(k):k{\neq}v\}$. Thus (4.10) implies that $\lim_v P[F(\mathcal{U}(v^*)) = t'] = 1$, which in turn implies that $\{L(\mathcal{U}(v^*),f)\}$ converges in probability to 0. This (in view of equation (4.6)) concludes the proof.

We proceed now to the demonstration of (4.10).

$$P[B(v) \mid A(v) \cup B(v)] \leq P[B_1(v) \mid A(v) \cup B(v)]$$

$$= P[t' > f(\mathcal{U}(v)) \geq t' - a(v)]/P[t' \geq f(\mathcal{U}(v)) \geq t' - a(v)].$$

As $\{a(v)\}$ converges to 0 monotonically, by the continuity property of measures,

$$\lim_v P[t' > f||(\mathcal{U}(v)) \text{ J } t' - a(v)] = 0.$$

Similarly,

$$\lim_v P[t' \geq f(\mathcal{U}(v)) \geq t' - a(v)] = P[f(\mathcal{U}(v)) = t'] > 0$$

Thus $P[B_1(v) \mid A(v) \cup B(v)] \to 0$, which in turn implies that $P[B(v) \mid A(v) \cup B(v)] \to 0$.

Corollary 4.3: Under the conditions of the theorem, one may compute a search procedure S_4 such that the results of the theorem still obtain, and further

$$L(X_1,f) \to 0.$$

(and consequently $f(X_1) \to \|f\|_P$) in P-probability for all $f \epsilon M$.

We describe the modifications of S_3' which achieve the result, leaving verification to the reader. For v=1,2,..., choose the K(v) sampling times randomly (i.e. uniformly) from M(v), M(v)+1,..., M(v+1) - 1. At the remaining times between M(v) and M(v+1), let X(t) = U((v-1)*). Observe that the sample times become sparse.

Proposition 4.4: Given positive numbers c,d, and e, and F_Z, the common distribution of the independent noise samples, there is a number N(c,d,e) such that for n > N(c,d,e), under the search described below, for all $f \epsilon M$,

$$P_f[P[f > f(X_n) + e] > c] < d.$$

Proof: {U(j)} is a sequence of N_1 independent, X-valued, P-distributed observations, where N_1 is a number large enough to assure (in accordance with Proposition 3.2) that

$$P[\min_{j \leq M} L(U(j),f) > c] < d/2.$$

Let h be a mapping with domain [-1,1] such that $h(a) = F_{Z+a}$. Then h is 1 to 1 and continuous with respect to the Prohorov metric (Prohorov [14] on the space of distribution functions. Consequently h^{-1} is uniformly continuous. δ is defined to be a modulus of continuity associated with e (assumed less than 1). N_2 is a number such that, in the notation of the Glivenko-Cantelli Theorem, for $n > N_2$,

$$P[\,|F - F_n|^* > \delta/2] < d/2N_1.$$

$N=N_1N_2$ and our search consists of sampling at each point U(j) and then letting N^* be the number j such that for some x and all k

$$F_{j,N}(x) > F_{k,N}(x) - \delta/2. \tag{4.11}$$

At times greater than N, the operating point is chosen to be U(N*). Toward showing that the strategy has the property given in the theorem, let U' denote the observation U(k) which minimizes L(U(j),f). With probability greater than 1 - d, simultaneously

(i) $L(U',f) < d$

and for $1 \leq j \leq N_1$,

(ii) $\| F_{j>N} - F_{Z+f\ (U(j))} \|^* < \delta/2.$

Assuming (i) and (ii) hold, by the triangle inequality and rudimentary properties of the translation parameter family F_{Z+a}, we see that (4.11) implies

$$\| F_{Z+f(U(N^*))} - F_{Z+f(u')} \|^*$$

$$= \| F_Z - F_{Z+(f(U')-f(U^*))} \| < \delta,$$

which, in view of the fact that the sup norm majorizes the Prohorov metric and also the way δ is defined, implies $f(U') < f(U(N^*)) + e$. This completes the proof.

We offer below some further refinements in the noisy measurement case. The proofs are only sketched as the ideas are similar to proofs already used.

Let $O_n = f(X_n) + Z_n$, the n<u>th</u> observed value. By $\hat{f}(n)$ we will denote any estimate of $\|f\|_P$. That is, $\hat{f}(n)$ is a measurable function of

$(X_1, \ldots, (X_n, O_1, \ldots, O_n)$. The basic idea in the next two results is the standard one of replicating observations to minimize the effects of observational error (see Brooks [9]).

Theorem 4.5: Let the Z_n be i.i.d.r.v.'s with a known distribution function F_Z. One may compute a search procedure S and estimates $\hat{f}(n)$ such that

1) $f(n) \to \|f\|_p$ a.s. (P_f) for all $f \in M$.

2) $L(X_n, f) \to 0$ i.p. (P_f) for all $f \in M$.

3) $\sum_{i=1}^{n} L(X_i, f)/n \to 0$ a.s. (P_f) for all $f \in M$.

Sketch of the proof: Pick Z a unique pth percentile for F_Z, that is $F_Z(Z) \geq p$ and $F_Z(Z<1) \leq p$ (for some fixed p, $0<p<1$). At any particular x by using the Kolmogorov-Smirnov approach and observing the pth percentile of $f(x) + Z_n$, $n=1, \ldots, N(\varepsilon)$ we may estimate $f(x)$ by the pth percentile $\hat{f}$ in such a way that

$$F_f(|\hat{f} - f(x)| < \varepsilon) \leq 1 - \varepsilon$$

Let $Y_1, Y_2, \ldots$ be i.i.d. X valued r.v.'s with distribution P. We proceed in iterations as in earlier theorems. During the nth iteration we have estimates $\hat{f}_1, \ldots, \hat{f}_n$ of $f(Y_1), \ldots, f(Y_n)$, respectively, all within ε_n with probability $> 1 - \varepsilon_n$. During the n+1st iteration most observations are at a point among $\{Y_1, \ldots, Y_n\}$ picked at random from among those Y_i satisfying $\hat{f}_i > \max\{\hat{f}_1, \ldots, \hat{f}_n\} - \varepsilon_n$. The other observations give estimates of $f(Y_1), \ldots, f(Y_n), f(Y_{n+1})$ to within ε_{n+1} with probability $> 1 - \varepsilon_{n+1}$.

During any given iteration, f(m) the estimate of $\|f\|_P$ is the maximum of the $\hat{f}_i$ of the previous iteration.

By choosing the ε_n such that the Borel-Cantelli lemma holds and by choosing (during each interation) a smaller and smaller fraction of X_i's to be used in estimating the $f(X_j)$'s one can show that the results of the theorem hold.

Theorem 4.6: For all $X \varepsilon X$ suppose that the noise $Z(x,f(x))$ satisfies one of the following:

a.) $Z(x,f(x))$ has a $n(\mu, {}^2(x,f(x)))$ distribution, μ known $\sigma^2(x,f(x))$ unknown.

b.) $Z(x,f(x))$ has a distribution which is symmetric about the known unique median μ.

c.) $Z(x,f(x))$ has a known mean μ and variance bounded uniformly above.

Then one may find a search procedure and estimates f(n) such that 1), 2) and 3) of Theorem 4.5 hold.

Sketch of proof: The thing to note is that in a), b) and c) for a fixed point x, ε, X if we repeatedly sample $X + W_i$, W_i independent and identically distributed with distribution the same as $Z(x,f(x))$ then for each $\varepsilon > 0$ we can find a stoping rule T and T-measureable estimates $\hat{f}(x)$ of $f(x)$ such that

$$P_f(|\hat{f}(x)-f(x)| \leq \varepsilon) \geq 1-\varepsilon.$$

In a) one could use the t-variable, in b) use the ideas expressed in Kendall and Stuart, [15] pages 513-522 and use Chebyshev's inequality in case c).

We thus can apply the same ideas as in Theorem 4.5.

If μ were unknown it would still be possible to satisfy conditions 2) and 3), but not condition 1.

We close this section by mentioning related studies. We have stated that in Kushner theory [7] it is supposed that f is a sample function of a known Brownian motion process. If is further allowed that the measurement may be corrupted by Gaussian noise having zero mean and a known variance, which is allowed to depend on the operating point x. The framework for computing an optimal search procedure minimizing $E[(\|f\|- f(x_n))^2]$ is sketched, but it is not proven that these methods yield convergence of the above expectation to 0.

Our studies are also somewhat related to the subject of "stochastic approximation," initiated by Monro and Robbins [16] and placed in an optimization setting by Kiefer and Wolfowitz [17]. A definitive survey of stochastic approximation has been written by Schmetterer [18]. Briefly, the stochastic approximation problem in determining the maximum of a regression function may be viewed as the problem of finding a search procedure yielding a sequence$\{X_i\}$converging (either in probability or or almost surely) to x*, where x* is the unique operating point maximizing f. The stochastic approximation setting is more general than ours in that the noise process, while (as in our studies) being independent of earlier observations, may be unknown and yet depend on x. But it is at the same time more restrictive than our theory because f must be a function which is unimodal. There are various other assumptions imposed on both F and the noise process; the reader is invited to consult the stochastic approximation literature.

5. COMMENTS

The goal in this paper has been to delimit what can be done by sequential search procedures when the set of objective functions is rich enough to include all continuous functions. This goal is more in the tradition of automata theory than numerical analysis. Where possible, we have sought bounds to the number of observations needed to accomplish those results that can be accomplished. Toward this goal we have revealed several search procedures giving convergence (in various senses) to optimal performance. Many of these results, especially in the noisy measurement case, are believed to be new.

For particular numerical problems wherein some prior knowledge of the criterion function f is available, we expect that often heuristic considerations will yield more rapid convergence than our algorithms. The literature suggests that heuristic "creeping search" programs (e.g. Schumer and Steiglitz [19]) have been used for some time. In any event, in computation, once the designer has found the number of searches, N, required to satisfy his tolerance of error, if the criterion function possesses any regularity whatsoever, it would seem sensible to sample at evenly spaced grid points rather than randomly chosen points as per the preceding algorithms. We suspect that the procedures we have proposed may have merit if the function f is easily evaluated (such as in linear or quadratic programming problems, etc.). Regardless of its computational merits (or lack thereof), the preceding analysis should have practical value in pointing out that certain search problems which are much more difficult than those currently studied are, in principle at least, amenable to solution.

Our viewpoint and procedures differ from other approaches to the sequential search problem in that the nature of the domain space can be suppressed. As noted above, the dimension of X plays little role, and in contrast with any other studies, the closeness of the operating point x to an optimizing point x* is of no consequence; it is on the closeness of f(x) to f(x*) that our attention focuses.

REFERENCES

[1] G. Hadley, "Nonlinear and Dynamic Programming," Addison Wesley, Reading, Mass., 1964.

[2] H.A. Spang III, A review of minimization techniques for nonlinear functions, SIAM Review, 4, (1962), pp. 343-365.

[3] J. Kowalik and M. Osborne, "Methods for Unconstrained Optimization Problems," Elsevier, New York, N.Y., 1969.

[4] J. Kiefer, Sequential minimax search for a maximum, Proc. Amer. Math. Soc., 4, (1953), pp. 503-506.

[5] J. Kiefer, Optimum sequential search and approximation under minimum regularity assumptions, SIAM Journal, 5, (1957, pp. 105-136.

[6] R. Bellman and S. Dreyfus, "Applied Dynamic Programming," Princeton University Press, Princeton, N.J., 1962.

[7] H. Kushner, A versatile stochastic model of a function of unknown and time-varying form, J. Math. Anal. and Appl., 5, (1962), pp. 150-167.

[8] H. Kushner, A new method for locating the maximum point in an arbitrary multipeak curve in the presence of noise, ASME J. Basic Engr. 86, (1964), pp. 97-106.

[9] S. Brooks, A discussion of random methods for seeking maximum, J. Opers. Res. Soc. of Amer., 6, (1958), pp. 244-251.

[10] A. Taylor, "Introduction to Functional Analysis," Wiley, New York, N.Y., 1958.

[11] P. Billingsley, "Convergence of Probability Measures," Wiley, New York, N.Y., 1968.

[12] K. Parasarthy, "Probability Measures on Netric Spaces," Academic Press, New York, N.Y., 1967.

[13] F. Massey, A note on the estimation of a distribution function by confidence limits, Ann. Math. Statist., 22, (1950), pp. 116-119.

[14] Yu. Prohorov, Convergence of Random Processes and Limit Theorems, Theo. Probability Appl., 1, (1956), pp. 157-214.

[15] M.. Kendall and A. Stewart, "Advanced Theory of Statistics," Vol. II, 3rd ed., Hafner Publ. Co., New York, N.Y. 1967.

[16] H. Robbins and S. Monro, A stochastic approximation method, Ann. Math. Statist., 22, (1951), pp. 400-407.

[17] J. Kiefer and J. Wolfowitz, Stochastic estimation of the maximum of a regression function, Ann. Math. Statist., 23, (1952), pp. 462-466.

[18] L. Schmetterer, Stochastic approximation, in "Fourth Berkely Symposium on Probability and Statistics," Vol. I, University of California, Press, Berkeley, California, (1961), pp. 587-609.

[19] M. Schumer and K. Steiglitz, Adaptive step size random search, IEEE Trans. Auto. Control, 13, (1968), pp. 270-276.